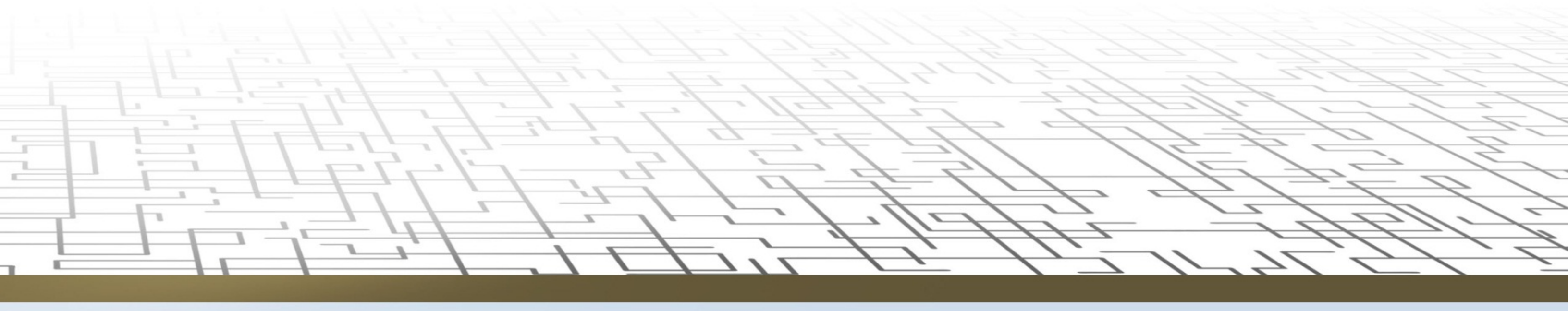


INFORMATION SHARING ENVIRONMENT

DATA AGGREGATION REFERENCE ARCHITECTURE (DARA)

NATIONAL SECURITY THROUGH RESPONSIBLE INFORMATION SHARING

PREPARED BY THE
PROGRAM MANAGER, INFORMATION SHARING ENVIRONMENT

VERSION 1.0

DECEMBER 2014

This page intentionally blank.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 EXECUTIVE SUMMARY

The need for greater interoperability is clear. To protect national interests, intelligence and law enforcement agencies must be able to collect, accurately aggregate, and share real-time analytical information about people, places, and events in a manner that also protects privacy, civil rights, and civil liberties. The President's National Strategy for Information Sharing and Safeguarding (NSISS) recognizes this as a priority national security issue, and speaks directly to this challenge. The Data Aggregation Reference Architecture (DARA) is in direct response to NSISS Priority Objective 10, "Develop a reference architecture to support a consistent approach to data discovery and entity resolution and data correlation across disparate datasets," The DARA provides a reference architecture that can enable rapid information sharing, particularly for correlated data, but also for raw data, by providing a framework for interoperability between systems, applications and organizations.

System owners throughout government would benefit from using DARA to design and implement system changes to enable greater interoperability. The DARA includes a Maturity Matrix (Appendix B), which assesses an organization's system maturity from Ad Hoc to Optimized across seven functional areas. DARA Version 1.0 has an initial Maturity Level target of Level 3. The target maturity level that will be achieved, with complete DARA implementation, is Level 5. The entire Data Aggregation Maturity Matrix is published now so that organizations may plan farther in advance for system lifecycle enhancements, and for building flexible and extensible systems, that enable progression to higher maturity levels. As organizations invest in system enhancements, cost effectiveness and return on investment may be measured in terms of system performance, mission capability, and greater national security.

As organizations continue these changes, leading to DARA implementation in larger numbers of organizations, greater interoperability occurs across the whole of government. As organizations and the community progress through maturity levels described in the DARA maturity matrix, they adopt community standards that lead to still greater interoperability. Over time, a higher degree of interoperability is achieved and organizations may quickly share information about correlated data and raw data with substantially less time required for system planning, development, and implementation than what would occur on a system by system basis once participating organizations have attained Level 5. The result is interoperability to share information at mission speed. Earlier maturity levels provide less external interoperability, but provide more consistent expectations around a known framework and shared lexicon for interoperability.

1.1 CURRENT STATE

Our current technical capacity to share information is uneven and particularly limited at the whole of government level. For example, in the current environment, an engineer at one organization may receive a bulk data transfer from another organization on a periodic basis. The bulk data transfer requires significant bandwidth, time to set up, and, inevitably, some time to troubleshoot. Once the data is transferred, the engineer hands the data off to a developer who performs additional processing on the data prior to making it available for analysts for queries and investigations. Data may become "stale" due to the delay inherent in periodic data transfers. Additionally, the time necessary for scheduling, troubleshooting, and processing decreased the data's timeliness and value.

This version of the DARA is designed to provide guidance with sufficient detail to guide readers towards development of program plans without prescribing implementation-specific requirements. It does not provide mandatory standards at this time, but enables organizations to develop, modernize, or modify systems to provide a basic level of interoperability with other participating organizations and their systems.

It provides a representation of a data correlation system that is expressed in terms of **Functional Areas** for data correlation: **Data**, **Structural Metadata**, **Discovery, Access Control**, **Change Data Management**, **Transport/Infrastructure**, and **Scalability**. The DARA is designed as an instructive guide for the three primary stakeholders (e.g., Executives, Program Managers, and Solution Architects). It provides a practical technical approach for the responsible assessment, planning, design/development, and implementation of an interoperable data aggregation investment.

1.2 FUTURE STATE

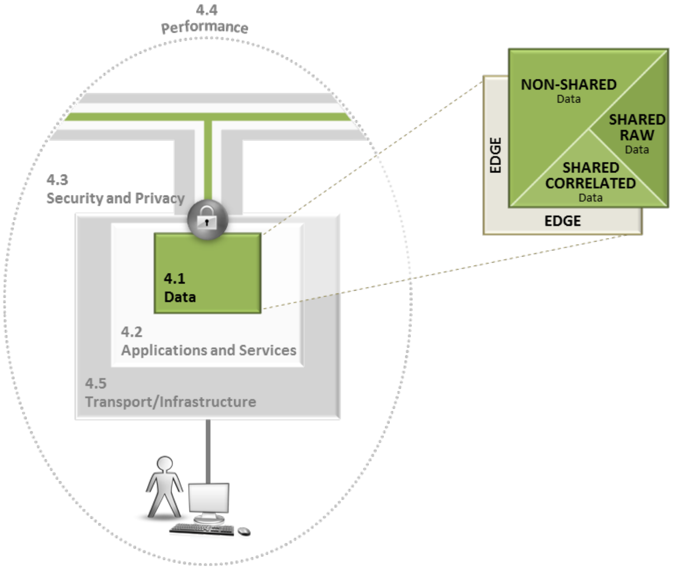
The ISA IPC Data Aggregation Report, released in 2012—*ISE Data Aggregation Capabilities Applicable to Terrorism*[1]—presented the findings and recommendations of the interagency Data Aggregation Working Group regarding the current state and potential futures of data aggregation efforts across government agencies both within and outside the Intelligence Community. It outlines three themes for improvement for data aggregation systems; The Need for an Improved Inter-agency Governance Framework; The Need for Improved Processes for Inter-agency Data Sharing Agreements and; accelerates the convergence of existing Data Aggregation Architectures and encourages development of Data Aggregation Reference Architecture (DARA).

The future state envisioned with full DARA implementation is the availability to participating organizations of raw and correlated data at the speed necessary to identify and counter rapidly evolving threats. This end-state will be achieved more quickly through the broad or complete adoption of community-wide standards as organizations implement the DARA, and the DARA and community systems continue to evolve over the next several years. An inability to share information at mission speed allows threats to unnecessarily evolve and inhibits identification, assessment, and response.

[1] https://max.omb.gov/community/download/attachments/736986154/2012-0518+ISE+Data+Agg+Capabilities+Report.pdf

Organizations that interoperate through the DARA improve analytical capabilities by increased accessibility of information contained in other organizations' systems. When analysts are able to search correlated data that other agencies provide using the DARA framework, then organizations do not need to replicate that information between systems, which saves storage space, bandwidth, and technical staff time across the entire federal enterprise.

The intent is to build a data and information continuum that requires agency level producers and aggregators to stage and expose their applicable information assets (i.e., leveraging already correlated (staged) data, and capabilities required for interoperability) by making data and information discoverable and sharable by other D/A consumers. Common services and other capabilities (ex., PII) that are required for interoperability will be developed and utilized by data consumers and providers.[2]



As departments and agencies implement changes to the agency's capabilities and bring information to the "edge", the interagency information sharing landscape changes to improve the mission value of the government's data holdings.

In an interagency data aggregation system assessment, it was determined that systems perform various levels of correlated data services based on the individual mission needs. In the context of the DARA, identified were three (3) types of data stores:

1) Non-shared Data – Data not for sharing outside of the organization
2) Shared Raw Data – Data dumps of basic transactional, uncorrelated data that is made available for information sharing
3) Shared Correlated Data – Correlated data that is made available for information sharing

Ultimately, an optimized data aggregation system will present fully correlated entity maps formatted using open standards with granular, attribute based access controls using "data tags" (resource attributes) that are generated with only limited manual intervention. Data exchanges operating at the higher maturity levels defined in Section 4 result in higher quality analysis and a faster "speed to intelligence."

[2] https://max.omb.gov/community/download/attachments/736986154/Data_Aggregation_Way_Forward_PO10+v1+05312013.pdf

1.3 AUTHORITIES FOR THE DARA

National Strategy for Information Sharing and Safeguarding[3] (NSISS) (December 2012) Goal 2.4 – Enhance Enterprise-Wide Data Correlation (Whole-of-Government).[4]

Strategic Implementation Plan of NSISS Priority Objective 10[5] implementation will "develop a Data Aggregation reference architecture to support a consistent approach to data discovery and correlation across disparate datasets."

ISE Data Aggregation Capabilities Applicable to Terrorism[6]*.* The ISA IPC Data Aggregation Report, released in 2012 *developed by the Data Aggregation Working Group*.

1.4 DATA AGGREGATION REFERENCE ARCHITECTURE DOCUMENT TOPICS

Enabling interoperability between agencies and departments requires guidance and standards in several areas. The DARA includes guidance for standards and actions in several domains:

- **Business**: The DARA includes a description of business domain maturity in functional areas that closely reflect the Federal Enterprise Architecture, focused on business process development, documentation, and change management that impact an organization's ability to manage a data aggregation system and interoperate with other organizations.
- **Data**: The DARA facilitates sharing of both raw (basic transactional) and correlated data. As the standards and practices in this reference architecture are adopted, data exchanges move from raw data to tagged, compound entity exchanges providing more value and more accurate data analysis.
- **Applications and Services**: The DARA includes a summary of core services that enable and implement the Data Aggregation Life Cycle. The DARA does not prescribe or mandate particular services architecture such as web services or the more robust Service Oriented Architecture. The definition of these services may vary within particular communities, or within a given enterprise architecture.
- **Security and Privacy**: Data handled by various governmental authorities is subject to differing legal and policy considerations regarding operational security, as well as the privacy, civil rights, and civil liberties of individuals and organizations described by the data. As such, organizations will likely caveat source data with various legal and policy

[3] http://www.whitehouse.gov/sites/default/files/docs/2012sharingstrategy_1.pdf

[4] Goal 2.4 – Connecting related information from disparate department and agency databases can mean the difference between identifying a threat during the planning stage or analyzing what could have been done to thwart the attack after it occurs. … To advance this capability while taking into account increasing volumes of information, stakeholders need to make their information accessible so an analyst can create a single query to search across many information sources …

[5] http://ise.gov/strategic-implementation-plan/po10

[6] https://max.omb.gov/community/download/attachments/736986154/2012-0518+ISE+Data+Agg+Capabilities+Report.pdf

access restrictions, and any operations on the source data will need to appropriately propagate those access restrictions to the resulting entity maps.

- **Performance**: Performance, in terms of this reference architecture for data aggregation and correlation, is broken out into two separate, but related areas: mission enablement and architecture and interoperability implementation. Architectural focus on performance both improves exchange speed and increases enterprise-wide scalability.
- **Transport/Infrastructure**: Organizations participating in information sharing possess rich IT resources to draw from and build on. In many cases, applying that infrastructure to the functionality covered in this document will be a refinement, not major development. In general, the Infrastructure reference architecture should be 'lightweight' with only the exposed interface documented with specificity. Internal infrastructure and methods do not need to be documented unless they affect the interfaces or access to the exchanged data.

1.5 DATA AGGREGATION IMPROVEMENT PROCESS

The DARA includes, in Appendix B, a Maturity Self-Assessment process which organizations should employ to assess their own maturity. Completing the assessment defines a system's current maturity state and provides organizations with the information they need in order to plan to achieve the initial target of Maturity Level 3. From there, organizations may plan for the actions and activities that are needed for a system to move to Maturity Level 5, which should include architecture development and capital investment planning.

2 ABOUT THIS DOCUMENT

Broadly, the DARA includes sections that address, in order:

- Introduction to the DARA, including key definitions and a description of purpose
- The Data Aggregation Reference Architecture, with specific sections for the following architecture domains:
  - Business Domain
  - Data Domain
  - Applications / Services Domain
  - Security Domain
  - Performance Domain
  - Infrastructure Domain
- A Data Aggregation Improvement Process that enables organizations to evaluate actions required to advance in maturity levels according to the Maturity Matrix
- References to other pertinent documents

Detail or specific descriptions are included in the following appendices:

A	Agency Data Aggregation System Profile
B	Maturity Self-Assessment
C	Relationships to Agency Standards
D	Service Mapping to the IC JARM (FOUO Version Only)
E	Sample Mission Use Cases
F	Glossary of Terms
G	Acronyms and Abbreviations

The DARA domains generally align with Federal Enterprise Architecture domains while the Maturity Matrix Functional Areas organize system characteristics more applicable to data aggregation and interoperability. The following table outlines the relationship between the DARA domains and the Maturity Matrix (Appendix B) functional areas.

Table 1. Relationship between the DARA Domains and the Maturity Matrix Functional Areas

DARA DOMAINS	MATURITY MATRIX FUNCTIONAL AREAS (APPENDIX B)
4.1 - Business	Cross-cuts all functional areas
4.2 – Data	Data, Structural Metadata, Change Data Management
4.3 – Applications / Services	Discovery, Data, Structural Metadata
4.4 – Security	Access Control, Discovery, Structural Metadata
4.5 – Performance	Scalability
4.6 –Infrastructure	Transport/Infrastructure

2.1 CONTRIBUTING ISE DEPARTMENTS AND AGENCIES

The Data Aggregation Working Group (DAWG) members were challenged to develop the future vision for data aggregation capabilities across the USG and establish a structural foundation for success. Realizing the major changes this entails, the DAWG broke the process down into near term and long term efforts with the intent of making incremental improvements for advancing large-scale data aggregation capabilities.

Data, Information, and System architects from departments and agencies (Table 2) are responsible for the development of the way forward vision, reference architectures, and establishing interagency partnerships to improve responsible information sharing and deliver mission capabilities.

In their capacity as co-chairs of the DAWG, technical representatives from the Office of the Director of National Intelligence’s (ODNI), National Counterterrorism Center (NCTC) and the Department of Homeland Security (DHS) provided executive leadership and organizational perspectives that greatly enhanced the utility of this document. **Special thanks** go out to the individuals for their contributions and their management for their time and talents.

Table 2. DAWG Participants Contributing to the DARA

Department of Homeland Security (DHS)
National Counterterrorism Center (NCTC)
Federal Bureau of Investigation (FBI)
Intelligence Community Chief Information Officer (IC CIO)
Office of the Program Manager – Information Sharing Environment (PM-ISE)
Regional Information Sharing Systems (RISS)
Department of State (DoS)
National Institute of Standards and Technology (NIST)

2.2 VERSION HISTORY

Version	Date	Comments
.044	23June2014	Release for Comments (1st Iteration)
0.5	23Sept2014	Release for Comments (2nd Iteration)
1.0	03Dec2014	Final

2.3 MAINTENANCE AND UPDATES

As the DAWG formed a working group to develop the DARA to achieve NSISS Priority Objective 10, the DAWG will continue to monitor the information sharing technology and policy environment within the United States Government and create a review and update path. Stewardship over the document may change over time, and the working group anticipates that updates to the DARA will occur when:

- Major new policies or legislation that impact the ability for or the governance of information sharing are implemented or enacted
- Major new initiatives, from the White House or between executive branch Departments and Agencies, occur
- Technology advances to the extent that current information sharing, security, or privacy and civil rights/civil liberties capabilities and practices no longer apply

3 INTRODUCTION

3.1 DEFINITION

Aggregation is the collection from across multiple sources. For data, it includes the collection of processes, policies, procedures, and technologies and linking information across organizations to allow for the detection of relationships between people, places, events, and other characteristics that are related to or an attribute of an entity that aid in establishing a contextual setting. The Data Aggregation Reference Architecture (DARA) is a technical reference architecture for primarily sharing correlated data, but also for improving the sharing of raw data when appropriately required by the mission. The purpose is to enable aggregation and integration of data for information sharing across agencies. It will describe the target state for all of government and provide guidance for agency aggregation systems to prepare them to interoperate with others.

In general, the DARA will:

- Provide an approach to assess the current state and a guide on how to move to the target state
- Assist in defining the interoperability requirements for data aggregation enterprise investments
- Define a reference architecture that enables entity resolution, data correlation and disambiguation across multiple data aggregation investments
- Specify what individual departments or agencies (D/As) need to do to embrace a federated approach and possible enhancements to their investments
- Encourage the use of (or definition of) and evolution of standards
- Serve as a broad, general reference architecture that guides the creation of more specific, concrete solution architectures
- Define performance metrics for scale and performance
- Provide directions to identify what individual D/As need to do to embrace a federated approach, and assess the possible organizational impacts to D/As

This version of the DARA is designed to provide guidance with sufficient detail to guide readers towards development of program plans without prescribing implementation-specific requirements. It does not provide mandatory standards at this time, but enables organizations to develop, modernize, or modify systems to provide a basic level of interoperability with other participating organizations and their systems.

It provides a representation of a data correlation system that is expressed in terms of **Functional Areas** for data correlation: **Data**, **Structural Metadata**, **Discovery, Access Control**, **Change Data Management**, **Transport/Infrastructure**, and **Scalability**. The DARA is designed as an instructive guide for the three primary stakeholders (e.g., Executives, Program Managers, and Solution Architects). It provides a practical approach for the responsible assessment, planning, design/development, and implementation of an interoperable data aggregation investment.

3.2 DOCUMENT SCOPE

The DARA is in response to Priority Objective 10 within the NSISS. As such, the DARA aligns with the scope stated in the NSISS. Therefore, the DARA is a technical policy document designed to facilitate the development of standards for a common interagency architecture for sharing aggregated information. The DARA does not address the complex and novel legal and policy challenges posed by aggregating data collected under different authorities (e.g., Title 50 intelligence agencies, Title 18 law enforcement agencies, civilian agencies) for different purposes (e.g., national security, human resources, law enforcement).

3.3 FUTURE STATE AND IMPACT TO ORGANIZATIONS

Achieving mission-speed interoperability is required in order to effectively investigate and respond to rapidly evolving threats or scenarios. The future state envisioned with full DARA implementation over the next several years is raw and correlated data available such that it is accessible via the broad or complete adoption of community-wide standards. The goal is to enable the sharing of correlated data between agencies to enable rapid data discovery, analysis and response. While there still may be some requirement for manual review of correlated data, the higher level of maturity that the community gains through the application of DARA principles will significantly reduce the amount of time spent manually reviewing records. This vision assumes a continued refinement and evolution of the DARA and its implementation in community systems.

Currently we share multiple, redundant copies of data sets in a raw, uncorrelated form with multiple consumers (data aggregation systems), so that the matching of a “person” or other entity must be done using both manual and automated processes prior to any analysis. Without a common data aggregation reference architecture, and associated standards and vocabulary, these processes are time consuming, error-prone, and utilize different methodologies between agencies. After following the DARA’s guidance to achieve a higher level of technical maturity, each agency will:

- Provide fully correlated maps of the data, with attribute-level sourcing generated, but still requiring manual approval.
- The data will be tagged at the attribute-level with open metadata standards.

- For discovery, there will be advanced search with predictive and prescriptive guidance and attribute-highlighted Entity Map results, configurable to federate between systems using an open standard.
- There will be granular, attribute-based access control based on open standards for access rules and data tags (resource attributes), high flexibility in assigning user attributes, and automated security and auditing procedures.
- The change data management will provide automated, event-driven, real time replacement of changed attributes, with history retained.
- The transport infrastructure will be configurable to operate with any system using an open standard with entirely automated pushes and pulls.
- There will be fully automated support for any conceivable data usage volume and additional sources to provide scalability. Agency systems are expected to share their data as well as consuming data from other participants.
- Nimble interoperability is achieved with dramatically less time, and other resources, expended to achieve interoperability between systems as systems conform to standard architecture and organizations adopt a standard lexicon for data aggregation and system interoperability.

Creating a single nexus or hub to facilitate interoperability would lead to a number of challenges and inefficiencies. Along with creating a new system come responsibilities for operations, maintenance, future enhancements, and customer service in order to maintain interoperability with many other systems which have their own lifecycles. Rather than building a new system or hub to facilitate interoperability, the DARA enables secure, nimble interoperability for the discovery and consumption of correlated and other available data across multiple agency holdings, to deliver “speed to mission” through focus on a common data sharing framework of capabilities, architecture and standards. Broadly, the community creates this capability as organizations enhance individual systems and the combined effect of system changes and enhancements, guided by the DARA, lead to greater interoperability across the community and across government. Consequently, in order to take advantage of this capability, organizations have a responsibility to enhance or modify their systems to enable applicable and appropriate sharing of their data.

3.4 PARTICIPATING AGENCY MATURITY LEVELS

In order to participate in interoperability efforts that emerge based on the DARA, each agency must achieve a particular level of maturity in the Information Sharing Environment. The Data Aggregation Maturity Matrix was designed so the majority of these capabilities can be realized when participants achieve maturity level 3 in the functional areas. Organizations should be able to evaluate the costs and benefits of increasing their organization’s maturity level. This is further described under the reference architecture below and in the maturity matrix in Appendix B that provides descriptions of characteristics descriptive of level 1 (Ad Hoc), to level 5 (Optimized). Based on analysis provided by this working group, it has been determined that an agency system should be at level 3 (Enhanced) to participate.

3.5 INFORMATION SHARING AND DATA PRIVACY

It is important to note that, as interoperability evolves and leads to a mission-speed information sharing capability, laws and policies that govern standards for privacy and civil rights and civil liberties will always apply – as noted in the Scope (see Section 3.2), this is a technical reference architecture. The DARA does not imply that every organization shares all data at any time. Instead, the DARA provides for the standards and architecture mechanisms that provide for seamless and mission-speed interoperability when appropriate use is already determined by the data owner.

3.6 BENEFITS AND DESIRED OUTCOME

The desired outcome of the DARA is to enable secure exchange of correlated data and access to data across the Federal government. This will be provided by:

- Reducing system operational costs and improving timeliness of information sharing via automated correlation and disambiguation across multiple datasets while minimizing or reducing the movement or copying of data.
- Providing enhanced protection and normalization of data (via hashing, anonymization, and encryption). Data providers remain the stewards of their data. Therefore, data will no longer need to be replicated on multiple systems, and their data privacy and security requirements will remain intact.
- Improving interagency collective knowledge of entities (e.g., persons, locations, organizations, etc.) and ability to react to emerging threats or indicators.
- Making data, information or IT services visible, accessible, understandable and trustable, which can be accomplished either directly or indirectly by the original data produces or indirectly via a designated third party.
- Enabling more rapid sharing of relevant data when a new "hit" is discovered.
- Increasing data quality as a result of search queries and faster updates to source data than via the current processes that require periodic data transfer after data is collected.
- Lowering incremental costs for departments and agencies through the use of existing systems to perform data correlation.

As organizations increase their maturity level from level 3 (Enhanced) to level 4 (Managed), previously manual processes will become more automated and faster, simplifying workflow, increasing the speed of information sharing, and becoming more extensible. Moving to level 5 (Optimized) provides for fully automated sourcing, advanced searching, higher flexibility of access controls, and event-driven change data management.

Additional benefits are described as they apply to specific use cases in Appendix E.

3.7 CONOPS OVERVIEW

Each participating organization may structure their unshared data, unshared raw data, and unshared correlated data in their own way, within the confines of their organization's guidance and the prescribed elements of this document. However, the shared raw data and shared correlated data should be presented in a common way, conforming to a defined and published contract that includes defined requests and common responses for data.

The graphic in Figure 1 depicts the high-level concept of operations (CONOPS) for a whole-of-government approach to a decentralized data-sharing environment. Figure 1 illustrates the DARA domains and the anticipated relationship between systems that interoperate by following the DARA. The DARA supports the set of decentralized common services for each data aggregation system, as depicted in the CONOPS.

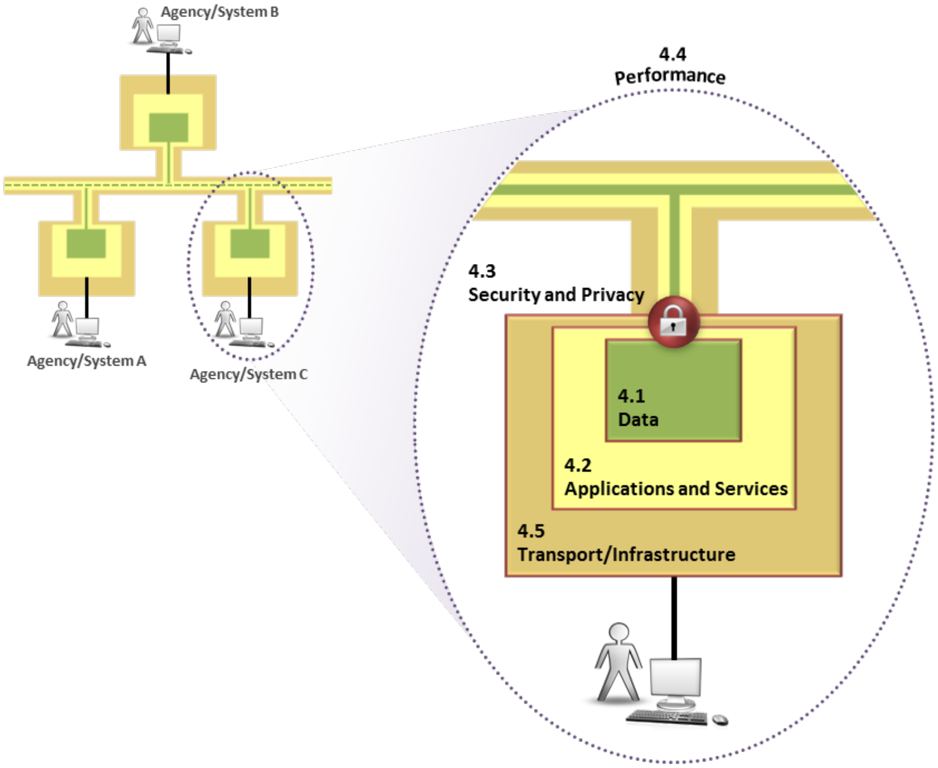


Figure 1. DARA Concept of Operations

Working group consensus indicates that an ideal data-sharing environment should include:

- **Distributed Data Sources** – Due to complications of ownership, change management, purpose, and use limitations, it is preferable that providers keep data under their operational control within the ownership of their originating organization, but make it available to consumers using standardized services and appropriate controls in a manner that, ultimately, creates a common framework for cross-system sharing of correlated data.
- **Correlated data** – To streamline typical data-related tasks, data collected from multiple sources should be linked and consolidated into entity records by each organization, each representing, for instance, a person, organization, event, or location.
- **Flexible metadata** – In order to allow for tightly selective transfer and presentation of data, along with information about security, lineage, provenance, pedigree, and legal and policy restrictions, the environment must allow for cell-level tagging of data. There is an existing dependency on NSISS Priority Objective 3[7] which will leverage existing standards in the development of the technical specifications where possible. DARA will incorporate the technical specification developed to support NSISS Priority Objective 3.
- **Data as a Service** – the resulting data, having been aggregated, standardized, and correlated by the originating organization, is available to participating organizations for their appropriate use through a standard set of interoperable services.

3.8 DOCUMENT APPROACH: HOW TO USE

The Data Aggregation Reference Architecture document is intended to assist in defining the interoperability requirements for data aggregation enterprise investments. New systems under development or existing systems can use this document to develop requirements or create a road map for interoperability enhancements. For best results, use the following 5 Step Approach in concert with the Maturity Self-Assessment in Appendix B to most effectively utilize the DARA, federal architecture frameworks, and other authoritative references throughout the document.

3.8.1 5-STEP APPROACH OVERVIEW

The 5-Step Approach below allows the most effective use of the DARA, primary architecture frameworks, and other authoritative references throughout the document. Use the ISE Information Interoperability Framework (I^2F)[8] Architecture Framework Alignment Grid, Data Aggregation Maturity Matrix, and Reference Architecture in the following steps. The steps are:

[7] Adopt metadata standards to facilitate federated discovery, access, correlation, and monitoring across Federal networks and security domains.

[8] http://www.ise.gov/sites/default/files/FINAL%20-%20ISE_I2F_v0%205.pdf

1. **Identify Mission Requirements** (Section 5.1.1) – Identification of authority(s), organizational requirements, discovery of system capabilities, data structures, and stakeholders using the System Profile Questionnaire (Appendix A).
2. **Perform Maturity Self-Assessment** (Section 5.1.2) –Consistently evaluate the maturation of the organization with regard to data, structural metadata, discovery, access controls, change data management, transport/infrastructure, and scalability using the Data Aggregation Maturity Matrix and Self-Assessment (Appendix B).
3. **Identify the Minimum Requirements for Interoperability** (Section 5.1.3) – Review self-assessment results to identify gaps in system capabilities for interoperability. Develop and recommend requirements to enhance capabilities for interoperability focusing on progress toward entity resolution, correlation and data aggregation goals in the areas of infrastructure, data, application and services, security, and performance.
4. **Use the DARA to Update Applicable Architecture** (Section 5.1.4) – Identify the desired interoperability level of each Functional Area of the DARA Maturity Model utilizing the Architecture Framework Alignment Grid.
5. **Build a Plan/Roadmap to Achieve Desired Interoperability Level** (Section 5.1.5) – Create an implementation plan to coordinate plans and track progress toward interoperability goals across departments and agencies.

3.9 DEPENDENCIES AND ASSUMPTIONS

3.9.1 DEPENDENCIES

Full DARA implementation and achieving Maturity Level 5 of the Data Aggregation Maturity Matrix depends on several parallel activities and community evolution. Generally, the working group is cognizant and acknowledges several categories of dependencies:

- Successful completion, by parallel working groups, of other Priority Objectives in the NSISS to resolve, for example, identity management issues that hamper or prevent cross-domain and cross-organizational access to information and that enhances discovery activities through a common standard implemented across domains and organizations. These Priority Objectives[9] are at various phases of their maturity and will be referenced in each section.

[9] Related Priority Objectives include:

- Governance (#1) – Align information sharing and safeguarding governance to foster better decision-making, performance, accountability and implementation of the Strategy's goals.
- Agreements (#2) – Develop guidelines for information sharing and safeguarding agreements to address common requirements, including privacy, civil rights, and civil liberties, while still allowing flexibility to meet mission needs.
- Data Tagging (#3) – Adopt metadata standards to facilitate federated discovery, access, correlation, and monitoring across Federal networks and security domains.

- Broadly, additional mechanisms and standards for identity management and metadata tagging.
- Legal and policy guidance that is necessary as the community moves to a model of mission-speed information sharing that addresses privacy, civil rights and civil liberties, acceptable use, technical architecture, and other operational and technical issues.
- A community-wide standard for sharing correlated data and correlated entity maps.
- Aligning information sharing and safeguarding governance to facilitate future decision-making and foster accountability, performance measurement, and implementation of future goals and standards.

3.9.2 ASSUMPTIONS

The DARA working group is well aware of these dependencies, and anticipates that organizations may implement temporary workarounds for many of these issues while remaining in compliance with legal and policy considerations. In order to facilitate these types of workarounds, the DARA includes the concept of a system profile, in Appendix A that organizations populate with information about their systems such that other organizations may review the information and gain sufficient knowledge to design and implement exchanges when their mission requirements would benefit from data in another system.

- The DARA team acknowledges that there are particular mission use cases requiring extreme analytic work that can only be performed in a central model, and the DARA framework supports this requirement by enhancing raw information sharing (as appropriate) in conjunction with correlated information sharing.
- Centralized and distributed models are complementary, with the salient advantage of the distributed model being the ability to make whole-of-government connections faster during an unfolding event (because of the availability of the entity maps/indexes and common data exchange protocols, and therefore simultaneously discoverable by multiple independent players).
- Partnering departments and agencies adopt a common information sharing agreement development process (refer to section 3.3) and engage the appropriate privacy, civil rights, and civil liberties professionals early in the process. This work is being implemented under the NSISS priority objective #2 to develop common guidelines for information sharing and safeguarding agreements.
- Partnering departments and agencies must work with appropriate privacy, civil rights and civil liberties professionals to develop guidance to address the potential impact on

- FICAM (#4) – Extend and implement the Federal Identify, Credential, and Access Management (FICAM) Roadmap across all security domains.
- Interoperability Baseline Capabilities (#6) – Define and adopt baseline capabilities and common requirements to enable data, service, and network interoperability.

individual rights and liberties created by new models of data aggregation and discovery that are not part of the traditional information sharing agreement development process.

4 REFERENCE ARCHITECTURE

4.1 BUSINESS DOMAIN (CROSS-CUTS FUNCTIONAL AREAS)

4.1.1 FUTURE STATE: BUSINESS DOMAIN

The DARA drives business domain activities, which influence all functional areas of the Data Aggregation Maturity Matrix, toward maturity as measured by ability to interoperate. The target state is for the community to define and document business reference models, business architecture, and business processes that:

- Accurately reflect an organization's needs for data.
- Provide for extensibility and scalability to incorporate new data technologies and to continue performing well under increased data loads and with increased users.
- Are documented using open standards that allow for rapid updates to schemas and architecture as requirements change.

4.1.2 SECTION CONTENTS

Activities within the Business Domain typically occur at the organizational level with operational concepts and capability descriptions that then drive system requirements. The functional areas within the Data Aggregation Maturity Matrix are directly applicable to systems which, in turn, have requirements that reflect the organization's business processes. As such, Business Domain activities, and descriptions below, generally align with the Federal Enterprise Architecture and may influence a system's maturity in all functional areas of the Data Aggregation Maturity Matrix. The maturity levels are described in the context of interoperability goals established by DARA. The DARA Business Domain addresses the following topics, in order:

- Business Reference Model
- Business Architecture
- Business Process
- Other Considerations

4.1.3 STAKEHOLDER PERFORMANCE GUIDE

Table 3. Stakeholder Performance Guide – Business

STAKEHOLDER PERFORMANCE GUIDE			
SECTION #1 –BUSINESS			
Role	Responsibility	Approach	Benefit
Executive Leadership	• Provide leadership support for transformation • Approval of investments in business process analysis and documentation • Approval for efforts to validate interoperability requirements in the organization's business context	• Identify the Policy, Governance, and applicable laws that direct the organization's mission • Oversee the documentation of organizational business processes, including flow-charts, sequence diagrams, stakeholder matrices, and other documentation • Identify and sponsor/champion the business and mission needs for interoperability to satisfy the mission requirements	• Provide important scope and contextual information for mission architecture and information sharing needs • Develop important insights into stakeholder positions and requirements • Inform strategic priorities for interoperability, such as interconnections with data providers, internal system enhancements, and others
Program Manager	• Define how DARA requirements are applicable to or fit into ongoing business processes for mission delivery and support systems • Identify opportunities for enhanced interoperability to further enhance business processes and resulting mission execution	• Incorporate business process development results that include requirements developed from DARA into program design and implementation plans • Drive stakeholders to define mission requirements for business process documentation and resulting interoperability requirements. • Assess the resource, scope, and implementation and work plans necessary to meet interoperability requirements as driven by business need • Communicate the scope, assumptions, and dependencies to Solution Architect	• Incremental and continuous movement towards the DARA, enabling identification of opportunities to share services government-wide • Documented business processes that lead to opportunities to reduce costs by eliminating duplicate functions or processes • Opportunity to increase buy-in, through documented business processes, on organizational policies, procedures, and functions across all stakeholders
Solution Architect	• Incorporate DARA features, capabilities and standards into solutions architecture as required by business process definition	• Incorporate DARA concepts and standards as they fit into business requirements applied to solution design and architectures • Assist in the development of the resource, scope, and implementation and work plans for achieve scope • Work with counterparts to define implementation details of interoperability when driven by business requirements	• Provides input technical integration and resource requirements to achieve scope of adoption of the reference architecture. • Provides technical expertise to development of application/service architecture and designs.

4.1.4 PARTICIPANT ROLES IN THE BUSINESS DOMAIN

The Data Aggregation community is comprised of multiple Departments, Agencies, Components and other organizations who each play multiple roles over time and in the context of their mission scenario. For the purposes of this document, we have identified two broad roles to clarify responsibilities:

- **Provider** – The role of the provider is the participant that makes the data available in correlated or raw form and understands and implements business architectures and processes that enable interoperability.
- **Consumer** – The role of the consumer is the participant that searches, discovers or receives data from another participant in the data aggregation community and implements business processes that enable stewardship and proper operational use of consumed data.

Table 4 summarizes the roles and responsibilities as related to business.

Table 4. Participant Roles and Responsibilities (Business)

DATA AGGREGATION ROLE	CORE DATA RESPONSIBILITIES (Mandatory)	TARGET STATE DATA RESPONSIBILITIES (Desired)
Provider	• Making key data holdings available and providing an API • Maintain agreed to or common data model/schema to ensure error free data exchanges	• Use open standards (e.g., Business Process Modeling and Notation (BPMN), Unified Modeling Language (UML)) to model business processes • Proactively provide information on updates to business processes when they affect data provided to other organizations
Consumer	• Validate business need for data provided by data providers • Update or develop business process documentation so that received data results in maximum operational value	• Use open standards (e.g., BPMN, UML) to model business processes, thereby increasing interoperability with Providers who do the same • Proactively stop using data from Providers when it is no longer needed so as to not use bandwidth or processing power that is not required by business need

4.1.5 BUSINESS REFERENCE MODEL

The Business Reference Model (BRM) provides organization scheme, or classification taxonomy, following terminology in the Federal Enterprise Architecture, for services performed in the Federal Government and distilled to those applicable to the organization participating in DARA for purposes of this document.

Level 1 – Limited or No Documentation or Organization of Business Functions. Organizations with limited or no documentation of business functions, or without change control and vetting processes, may only have limited insight into their business operations. Similarly, for an additional reference point, other documentation typically required by federal organizations, such as OMB Exhibit 300/53[10] materials or Congressional Justifications, can only be completed on an ad hoc basis. With respect to interoperability, organizations at Level 1 may not understand their data needs, interoperability capabilities, and have very limited capacity to interoperate with other organizations.

Level 2 – Documented and Organized Business Functions within an Organization. Organizations generally have completed documentation of business functions and may have limited configuration management and vetting processes for managing change. Organizations can generally complete OMB Exhibit 300/53 and similar federal requirements in a manner that is repeatable and consistent across their programs and documentation. For interoperability, organizations business function documentation generally provides good context for their systems' interoperability requirements, but address or develop interoperability requirements and capabilities on a case by case basis.

Level 3 – Documented and Organized Business Functions across the Enterprise. Organizations have complete documentation of business functions along with a standardized and repeatable change control process. Organizations document business functions in concert with and with relationships to enterprise business functions where appropriate (e.g., when an "organization" could be an agency that is part of a department). Similarly, processes to meet OMB and other federal requirements are organized, standardized, and repeatable with consistent reporting and findings across all documentation and programs. Organizations understand data needs in order to meeting their mission requirements and have incorporated interoperability functions into their business function documentation.

Level 4 and 5 – Documented and Organized Business Functions adhering to Community and Federal Standards. Organizations have complete documentation of business functions along with a standardized and repeatable change control process at the enterprise level. Business function documentation adheres to a community or federal standard and can be easily understood by other organizations. Processes to meet OMB and other federal requirements are organized, standardized, and repeatable with consistent reporting and findings across all documentation and programs. Organizations have incorporated interoperability functions into business function documentation and commit to using open standards to document associated business processes.

[10] http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/fy14_guidance_on_exhibits_53_and_300.pdf

4.1.6 BUSINESS ARCHITECTURE

Business architecture contains the detailed information that links the BRM to organizational goals, objectives, policies, organizational structure, business functions, processes, and policies to technical architecture, including Data, Applications/Services, Security and Privacy, Performance, and Transport/Infrastructure. The linkages or relationships documented in the Business Architecture enable better organizational assessments of the effectiveness of different activities that take place or changes that occur on actual mission results.

Level 1 – Business Architecture not Documented or Minimally Documented. Organizations have not documented or have minimal, potentially aged or inaccurate, documentation available to link business architecture with business functions. Capability for interoperability is limited or potentially does not exist. When organizations' systems interoperate with other organizations' systems, the results may be unpredictable or unreliable. Organizations at this level may not be able to show ROI from interoperability, nor estimate organizational changes or investment that may be required to interoperate.

Level 2 – Business Architecture Documented for the Organization. Organizations may have documented business architectures, but may not have had any third party review or validation and may not have change control processes in place. Stakeholder requirements and expectations within the organization are generally understood. There is little or no linking of organizational business architecture with enterprise business architecture for organizations that are a part of larger organizations (e.g., agencies that are a part of a department). Interoperability may occur at the system level, on a case by case basis, and could be reliable if organizations' business architecture documentation is governed by change control and vetting processes and with additional details for change management and notifications specified via MOU/MOA.

Level 3 – Business Architecture Documented for the Organization and the Enterprise with Associated Relationships. Organizations have documented business architectures with standardized and repeatable change control processes and with documented linkages and relationships to enterprise business architectures and stakeholder expectations. Interoperability with other organizations' systems is feasible, generally requiring minimal technical development. Interoperability is documented as a part of the business architecture with well-understood linkages to organization and enterprise mission requirements in a manner that enables rapid analysis of mission impacts should interoperability become further enhanced or degraded.

Level 4 and 5 – Business Architecture Documented to Include Interoperability Across the Community and Government. Organizations have documented business architectures with well-defined and understood linkages and relationships to enterprise, community, and government business architectures with well-understood stakeholder requirements and dependencies including for stakeholders outside of the organization. The requirements and impacts of interoperability are documented in business architecture and are well understood or easily estimated. Organizations can easily derive system requirements that depend upon business architecture documentation and estimate or predict the opportunities and impacts of further community interoperability enhancements. At the same time, organizations can identify redundant or duplicate functions within their enterprise or community in order to create opportunities for efficiency and saved resources.

4.1.7 BUSINESS PROCESS

Business Processes are documented and modeled to enable more thorough understanding of organizational business operations and therefore more accurate mission and system requirements. In some cases, and in organizations operating at higher maturity levels, business processes are documented using open standards that translate directly to application execution and data schemas.

Level 1 – Business Processes not Documented or Minimally Documented. Organizations have not or have only minimally documented business processes. System requirements that should depend on or reflect business processes are instead documented on an ad hoc basis without any corresponding linkages or relationships to organizational operations and stakeholder requirements. Organizations may only achieve interoperability capabilities on an ad hoc basis and with low expectations for stability, reliability, or future enhancements.

Level 2 – Business Processes Documented for the Organization. Organizations have documented business processes but may lack change control and vetting processes and the ability to validate updated system requirements to correspond to business process changes; business processes are primarily a snapshot in time. Organizations have a good understanding of stakeholder requirements and dependencies within their organization. Organizations achieve interoperability between their systems and other organizations' systems on a case by case basis without a good understanding of requirements and dependencies between organizations, resulting in interoperability that may not be stable or reliable.

Level 3 – Business Processes Documented for the Organization with Contextual Relationships to the Enterprise. Organizations have documented business processes with well-defined linkages and relationships to enterprise business processes and with standardized and repeatable change control and vetting processes. Stakeholder requirements and dependencies are documented and well understood, and organizations can easily understand or predict the impact on operations if they enhance or degrade interoperability capabilities. Organizations have documented business processes sufficiently to enable rapid translation to system requirements and to quickly estimate the impact of business process changes on system requirements or of the potential for new technologies to impact business processes.

Level 4 and 5 – Business Processes Documented to support Community Interoperability Using Open Standards. Organizations have documented business processes using open standards (e.g., BPMN, UML) with well-defined linkages and relationships to enterprise, community, and federal business processes. Stakeholder requirements and dependencies are documented and well understood, including requirements and dependencies from external stakeholders, and organizations can easily understand and predict the impact on their own or other organizations' operations if they enhance or degrade interoperability capabilities. Organizations have documented business processes that, in concert with open standards that enable rapid executable or schema generation, are easily adapted to changes in interoperability status or capabilities within an organization, the community, or federal government.

4.1.8 OTHER CONSIDERATIONS

4.1.8.1 MOU/MOA

Organizations that interoperate may require that specific expectations for data use, data attributes, and system performance be documented in an MOU/MOA. The DARA v1 begins to address interoperability by establishing a framework at a target maturity level that does not include the community-adopted or open standards that would be required at higher maturity levels. In the future, some community-adopted or open standards, when widely used at higher maturity levels, may speed establishment of MOU/MOAs between organizations. At this point, the DARA working group anticipates that some expectations for interoperability capabilities are normalized, but that organizations will execute MOU/MOAs for specific operational requirements.

4.1.8.2 VARYING FEDERAL ARCHITECTURE STANDARDS

The DARA working group recognizes that organizations' adherence to Business Domain levels of maturity described in this section may vary according to their departmental or community enterprise architectures that may differ from the Federal Enterprise Architecture. The DARA expresses maturity levels in general terms with the expectation that organizations can map the descriptions to their enterprise architecture frameworks.

4.2 DATA (FUNCTIONAL AREAS 2, 3, AND 6)

4.2.1 FUTURE STATE: DATA

The DARA facilitates sharing of both raw (basic transactional) and correlated data depending on the mission use case. The target state for the community is moving towards real-time, incremental, service-based exchanges, although bulk raw data transfers may still be required in select cases. As the standards and practices in this reference architecture are adopted, data exchanges move, from raw data, toward tagged, compound entity exchanges providing more value and more accurate data analysis. Ultimately, an optimized data aggregation ecosystem will present fully correlated maps formatted using open standards with granular, attribute-based access controls and data tags (resource attributes) generated with only limited manual intervention required, and then primarily when called for by system limitations, organizational procedures, or to meet legal and policy requirements. Data exchanges operating at the higher maturity levels defined in Section 3 result in higher quality analysis and a faster "speed to intelligence." Future versions of the DARA will begin to define specific standards for tagging and data sharing, based on Intelligence Community Information Technology Enterprise (IC ITE), National Information Exchange Model (NIEM) and other standards already in use. The goal is to harmonize and interoperate, not reinvent standards.

Note that another class of data is "unstructured" which may consist of free form text data or non-text data types. As participants mature their capabilities for data sharing, new concepts for unstructured data sharing will be incorporated into future versions of the DARA. As volumes of data increase, the use of automated correlation tools increase to assist in processing and analyzing data. Current and future systems balance automation with the need for human oversight to validate linkages and conclusions, especially in early stages of a new capability.

4.2.2 SECTION CONTENTS

While Data is in itself a Functional Area, this section touches a total of three functional areas from the Data Aggregation Maturity Matrix. Figure 2, below, shows Data with respect to other DARA Functional Areas and this section. It addresses these four areas as well as other relevant concepts in the following order:

- Correlated Entity Data Standards (Cross-cutting Functional Area)
- DARA Functional Areas
  - Data (Functional Area 2)
  - Structural Metadata (Functional Area 3)
  - Change Data Management (Functional Area 6)
  - Other Considerations

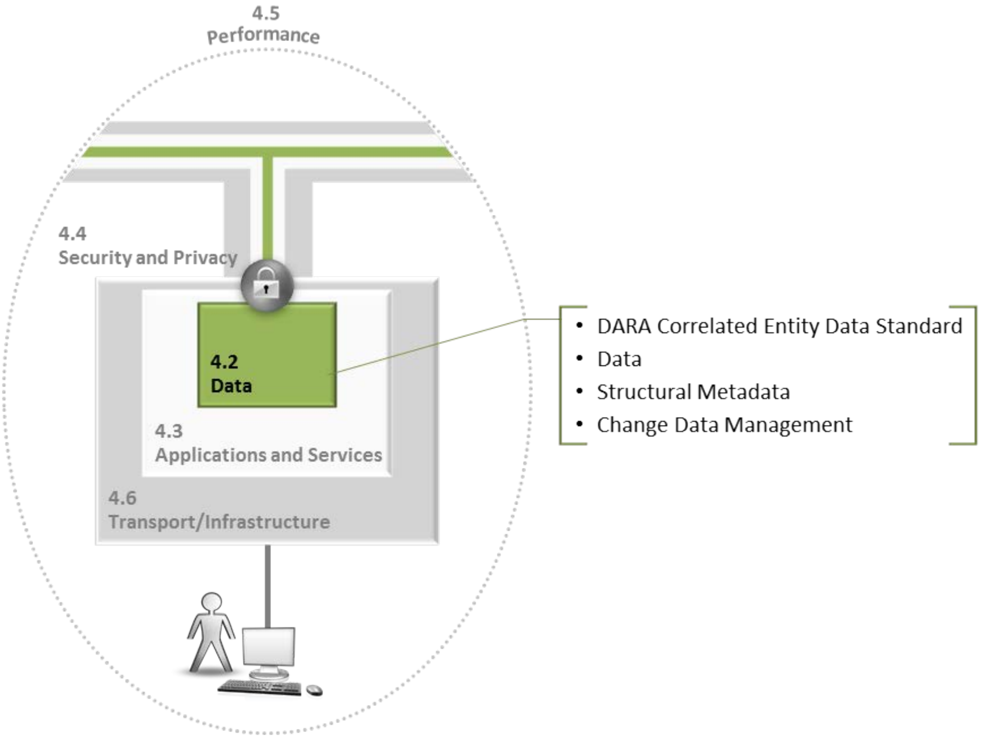


Figure 2. Data with respect to other DARA Functional Areas

4.2.3 STAKEHOLDER PERFORMANCE GUIDE

Table 5. Stakeholder Performance Guide - Data

STAKEHOLDER PERFORMANCE GUIDE			
SECTION – DATA			
Role	**Responsibility**	**Approach**	**Benefit**
Executive Leadership	• Develop an understanding of how data management changes under a whole of government approach • Input and approval on data access rules	• Define what Policy, Governance, and applicable laws that will affect the implementation of the architecture. • Develop guidelines for information sharing and safeguarding agreements to address requirements, including privacy, civil rights, and civil liberties. • Identify the business and mission needs for interoperability to satisfy the mission requirements. • Participate in "cross-agency" meetings to define standards needed to implement NSISS objectives and make the DARA a success • Develop oversight structures, rules and processes for addressing the potential impact to privacy, civil rights and civil liberties posed by new data aggregation and visualization	• Provide guidance, documenting boundaries and mission value to reference architecture. • Documents the mission needs for interoperability and information sharing. • Documents the roadmap/plan to maintain progress toward interoperability goals and to coordinate plans across departments and agencies.
Program Manager	• Define how DARA requirements are incorporated into program plans and delivery schedules	• Incorporate data architecture concepts and standards from DARA into program design and implementation plans. • Drive stakeholders to define mission requirements for data tagging, data access rules and interoperability requirements. • Assess the resource, scope, and implementation and work plans for achieve scope. • Communicate the scope, assumptions, and dependencies to Solution Architect.	• Incremental and continuous movement towards the DARA whole of government data sharing approach. • Documented work plan and roadmap that leads to interoperability. • Coordinate with all stakeholders the desired interoperability architecture artifacts descriptions and scope.
Solution Architect	• Incorporate DARA features, capabilities and standards into data architecture, data tagging and overall architecture	• Incorporate data architecture concepts and standards from DARA into solution design and architectures. • Assist in the development of the resources, scope, and implementation and work plans to achieve interoperability goals • Work with counterparts at Partner Agencies to define implementation details to ensure interoperability.	• Provides input technical integration and resource requirements to achieve scope of adoption of the reference architecture. • Provides technical expertise to development of data architecture and designs.

4.2.4 PARTICIPANT DATA ROLES

The Data Aggregation community is comprised of multiple Departments, Agencies, Components and other organizations who each play multiple roles over time and in the context of their mission scenario. For the purposes of this document, we have identified two broad roles to clarify responsibilities:

- **Provider** – The role of the provider is the participant that makes the data available in correlated or raw form.
- **Consumer** – The role of the consumer is the participant that searches, discovers or receives data from another participant in the data aggregation community.

Table 6 summarizes the roles and responsibilities as related to data.

Table 6. Participant Roles and Responsibilities (Data)

DATA AGGREGATION ROLE	CORE DATA RESPONSIBILITIES (Mandatory)	TARGET STATE DATA RESPONSIBILITIES (Desired)
Provider	• Making key data holdings available and providing an API • Tag data with standardized access and discovery tags • Maintain agreed to or common data model/schema to ensure error free data exchanges. • Provide access instructions, mission context, and business rules and policies • Receive data quality feedback	• Make correlated entity maps available according to the DARA standard • Receive back “enriched data” where applicable
Consumer	• Respect and enforce data access tags • Retrieve data within mission parameters and data sharing agreements • Understand and apply change data • Understand and apply Redress requirements (data or correlation corrections) • Meet the audit requirements specified by the provider for this data • Structure system calls by passing appropriate user attributes to authenticate requests • Preserve all metadata tags provided by the original provider	• Develop enriched data and share with original provider • Add new metadata tags for security, etc. when the data is changed • Provide feedback on Data Quality Issues • Add to the consumer responsibility

4.2.5 DARA CORRELATED ENTITY DATA STANDARD (CROSS CUTTING FUNCTIONAL AREA)

One goal of the DARA is to enable the sharing of Correlated Entities across organizational boundaries. Conveying the correlation method and context are key challenges. Sharing of correlated data and resolved entities builds on the same concepts as sharing raw data—standard formats, data tagging, retention, privacy—but adds complexity and nuance in each of those areas. This section defines additional data sharing guidance for sharing correlated entities. To address this gap, the DARA sets out to define a Correlated Entity Data Sharing Standard based on input and best practices from DARA participants and industry. The DARA working group issued an industry Request For Information (RFI) (routed through DHS S&T) to ascertain if an industry standard existed for sharing correlated data across organizational boundaries in way that conveys confidence in correlation out of the context of the original correlation system. The result of the research indicates that no single standard does exist for this, although industry provided several good avenues for further discussion.

A future version of the DARA (DARA 2.0) and related activities will solidify a working standard for the participants to use. In advance of that the following section defines the concepts that a correlated data sharing standard will address. A correlated entity map includes at least the following components:

- **Top-level records representing entities** (e.g., persons, locations, organizations, etc.) with attributes consolidated from multiple sources containing records on that entity.
- **Ontological information** showing relationships among top-level entities (i.e., person entity A is a member of organization entity B).
- **Summarized (or complete) source records** for reference in connection with top-level entity, including the primary attributes of the entity’s record in each source system and potentially a confidence score of each source entity record (e.g., collected from a passport scan vs. data entry on a website).
- **Source system metadata**, such as a real-time operational point of contact for verifying any information before taking related action.
- **Configuration information for record linkage** operations performed on source systems to create consolidated entities, i.e., what software, models or algorithms were used to form the linkage.
- **Computed confidence values** attached to attributes, records, and/or sources, for use in gauging trust in the defined linkages. This is the current gap in the standard, how to convey confidence values across organizations and across different correlation systems (whether COTS, GOTS or custom code) that the receiving consumer can understand without having to break apart records and “re-correlate”. Future versions of DARA will focus on closing this gap.

- **Data Assurance Information defining the accuracy, trust, precision and reliability** of data and systems. Feedback on system quality issues must be factored in overtime as quality and reliability of different systems, data sets or even attributes are determined.

4.2.6 DATA (MATRIX FUNCTIONAL AREA 2)

As the DARA community matures in data sharing capabilities, organizations when acting as providers advance their capabilities to share data in more complex and meaningful ways. Data structures carry correlated compound entities whose individual components might have different confidence or correlation factors and data tagging provides both constraints and context. Participants adopt effective policies to secure the privacy, civil rights, and civil liberties of individuals on whom data has been collected while allowing functional access to critical data by authorized personnel.

Level 1 – Raw (Transactional) Data Exchange. The first level of maturity in data sharing is providing a full copy of a source system with no or little manipulation. Raw data sharing implies a simple database copy, a limited export or other basic formats. Sharing full database copies of data is sometimes a mission requirement, but participants should seek to define an incremental, periodic transfer instead.

Level 2 and 3 – Basic and Enriched Entity Records. As providers offer data formats specifically designed for external consumption, they use common data structures and methods to define entities. A schema leverages some common formats. Data providers tag transactional data with pedigree, retention, and privacy indicators (as defined below). **Sharing well-formed, tagged data is the target maturity for DARA v1.0 participants.** Moving beyond the DARA v1.0 target of level 3 and DARA data providers develop more sophisticated mission systems internally, their capabilities to share data across the community mature as well. Data Aggregation participants in with advanced capabilities should push beyond the immediate target for more sophisticated sharing at Level 4 and potentially Level 5 for meaningful, mission relevant data sharing.

Level 4 – Partially Correlated Data for an Organization. An example of this is a “person-centric” data set where records (entities) are organized around a strong identifier such as passport number or other travel identifier. Parts of the data are correlated but with limitations.

Level 5 – Full Organizational Correlation. Level 5 of data sharing occurs when a data provider uses a sophisticated rules-based or probabilistic matching engine. The provider shares a complex, compound entity map using the **DARA Correlated Entity Data Format**, which will be defined in DARA v2.0 in accordance with concepts defined above. The data-sharing payload includes multiple entities each with their own data tagging but also information about the linkages. The linkages identify how the entities are related, the method of correlation and the relative strength of correlation in the originating aggregation system. The goal of sharing correlated data is to provide actionable data that is already processed, saving the consumer significant processing time. The information about the linkage provided context and confidence levels regarding the linkages so the consumer can act accordingly. Some consumers may use the linkages as provided, while other consumers may use the linkage as one data point and run additional correlations with their own data sets.

4.2.7 STRUCTURAL METADATA (MATRIX FUNCTIONAL AREA 3)

In order to accurately interpret and audit data, it is important to have a clear operating picture of the historical pedigree of the data. Data providers include metadata in a scaffolding that captures details such as time, authority, data retention and operation parameters in a manner that supports the other functional areas that may leverage it. Data consumers must honor and retain provided tags and potentially provide additional tags if the data is transformed, linked, or increased in sensitivity or classification, including as a result of aggregation. Data providers make data available (whether as an extract or in response to a service call) in structured well-formed data structures. Formatting is required at multiple levels within a data extract or message. An overall "envelope" container has tags to indicate the organization, extraction date, source system. The "payload" has more specific tagging on individual occurrences within the data for discovery and security controls.

Data tagging provides several functions, first and foremost to provide a foundation for controlling the data—what classification or sensitivity level, how long can it be kept, in what mission scenarios can it be used, the sensitivity of the data, including any legal or policy restrictions on its use, what conditions must exist? Tagging should define where data comes from (i.e., "Provenance") or "Pedigree." Tagging can also indicate the reliability of the data and whether it is derived from a biometric, a machine-readable document (e.g., a biometric, a machine-readable document such as a passport versus a website entry), or a website entry, or other sources. Finally, metadata to support Discovery indicates functional domains, specific identifiers and other descriptive identifiers.

Interpretation and use of the metadata is largely mission-dependent, although data retention and privacy tags must be enforced consistently. Legal and policy mandates for particular actions may require mission operators to purge data past a certain age with the system, or define new retention standards with the data owner. In the absence of clear guidance for any organization, developing such guidance should be encouraged.

Level 1 – Basic Data Tagging. Participants in DARA store data internally in ways that best meet their mission obligations. For exchanging data among DARA participants, a data provider should make data available in a tagged, structured format.

Levels 2 and 3 – Standards-Based XML Tagging. Participants use standardized XML data formats metadata tags, with preference given to agreed-upon metadata standards (such as the metadata guidance given by NIEM and IC-ITE). Systems have the capability to tag data, for access control purposes, prior to being shared externally. As long as an agency is providing well-formed and predictable data formats, even if in an agency-specific format, they are meeting the bar for Level 3. Level 3 is the target level of maturity of DARA v1.0 participants.

Levels 4 and 5 – Granular Attribute Based Data Tagging Based on Community Standards. Ultimately, an optimized data aggregation system will present data tagged at a granular level with metadata standards as defined in IC-ITE and other detailed data tagging schemes. As all participants come together to develop whole of government standards for person data, extended attributes and encounter data, participants can elevate to levels 4 and 5 Access Control (Matrix Functional Area 5). Future versions of DARA will explain and define the metadata and access tagging in more detail, mapping to existing standards or extending them where needed. The goal is to establish inter-operable tagging schemes vs. creating one master whole of government tagging standard.

To support standardized tagging, the DARA in future versions will identify existing standards or propose new standards for tag definitions. To maximize automated exchanges, tagging sets will need to be consistent or at least mappable to each other.

4.2.8 CHANGE DATA MANAGEMENT (MATRIX FUNCTIONAL AREA 6)

To make data sharing more effective and lower friction, data providers need mechanisms to indicate changes in data since that data was last provided to that particular consumer or changes since a particular frequency (e.g., changes since the last week or changes since the last month).

Change Data mechanisms start out straightforward but increase in complexity as the number of participants increase and as the variety and sophistication of the data payloads increase. From the simplest case, a monthly batch exchange process can tag records as being a Change, Add or Delete since the previous iteration of the batch export. However, if the data payload is a compound resolved entity requested on an ad-hoc basis, both the core entities might change and the correlation links or strength of links may change. Challenges arise as complex sharing is involved, and complex issues will need to be addressed. For example:

- A consumer agency’s need to know when a source agency implemented a redress process that results in unlinking of records.

- If a provider organization changes its correlation rules or data that impact previously communicated correlations, how the organization conveys that information.

Level 1 – Full Data Replacement – On a periodic basis, an entire data set is re-extracted and re-transmitted to the receiving entity. The full data set needs to be reprocessed and analytical work may need to be redone depending on the receiving system.

Level 2 – Simple Timestamp Driven Changes – On a periodic basis, a provider creates an extract of all data that has changed since the previous extract date. Data volumes are lower but the change data logic is fairly coarse operating at the system or data set level.

Level 3 – Automated Change Management with History – Change records are sent to the data consumer with a history of changes available, at least internally, if not externally. History provides an audit trail of data changes and maintains security tags in case there is a question about a change or a need to go back to previous values.

Level 4 – Event Driven Change Management – Functional or system events trigger change data to be recorded or sent to participating systems. This requires the combination of data and services to support, as described in the Applications / Services section for “Persistent Search and Alerts”.

Level 5 – Near Real Time. High priority changes are identified and transmitted to participating entities in near real time to support mission operations. Integrations of correlation systems at near real time causes some challenges to deconflict competing changes as systems get out of sync with each other.

4.2.9 OTHER CONSIDERATIONS

4.2.9.1 FREQUENCY OF DATA UPDATES

Consideration for latency falls primarily under the Change Data Management row of the maturity matrix. To avoid capturing actual mission needs or system-specific goals, the matrix only considers latency in the final column. Latency is irrelevant for levels 1–4, but near real time update capability is required for a system to be considered Level 5, or Optimized. This is a compromise that enforces real-time sharing as the ultimate goal for some mission systems, while allowing that its difficulties eliminate it as a short-term goal. The DARA and Maturity Matrix will continue to consider mission needs vs. technology and investments required to meet those needs.

4.2.9.2 DATA QUALITY AND FEEDBACK LOOP

This Data Aggregation Reference Architecture delineates that data consumers have obligations to the data providers regarding how data is ingested, transformed, and used or “enriched.” Data Quality issues should be proactively communicated back to the originating data set. Changes to the data could involve new correlations, additional data or discovery of false information. In the target state, enrichments to the data should be made available to the data set originators. In summary:

- Consumers should consistently provide feedback to providers regarding data standardization and data inconsistencies with the goal to increase the overall enterprise data quality standards for both sides.
- If allowable within the mission constraints, if the consumer enriches the data through additional attributes, linkages or other transformations, the DARA expects that the new data should be shared back to the original provider who may or may not have the ability and mission need to update the originating system, but nevertheless may have a use for the new information.

The Data Aggregation Working Group understands that each participant is already moving towards existing Federal, Agency and industry standards related to data sharing and data aggregation. The DARA attempts to harmonize these efforts along the lines of sharing correlated and other data types.

4.3 APPLICATIONS/SERVICES (FUNCTIONAL AREAS 2, 3, AND 4)

4.3.1 FUTURE STATE: APPLICATIONS AND SERVICES

The DARA facilitates the development and adoption of applications and services that interoperate by providing common definitions for services and terminology that enable understanding and comparison between applications and services used on different systems. The target state for the community is use of well-understood applications and services that enable clean, efficient interoperability of data, including functions for aggregation, correlation, resolution, and discovery. Following the Maturity Matrix, an optimized approach to applications and services enables seamless discovery, automated data attribute updates and enrichment, and data interchange between organizations.

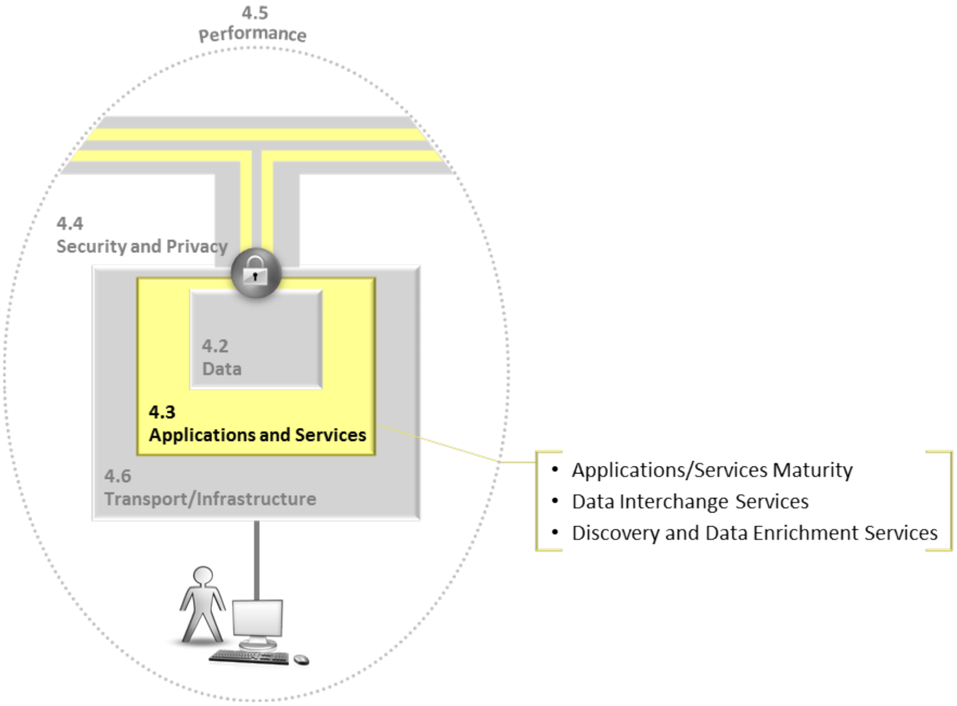

4.5
Performance
4.4
Security and Privacy
4.2
Data
4.3
Applications and Services
4.6
Transport/Infrastructure
• Applications/Services Maturity
• Data Interchange Services
• Discovery and Data Enrichment Services

Figure 3. Applications and Services with respect to other DARA Functional Areas

The following is a summary of core Services that that are included within this version of the DARA. The DARA does not prescribe or mandate particular services architecture such as web services or the more robust Service Oriented Architecture. The definition of these services may vary within particular communities, or within a given enterprise architecture.

4.3.2 SECTION CONTENTS

The Applications and Services section cuts across multiple functional areas from the Data Aggregation Maturity Matrix as it depends upon activities and technologies that implement multiple functional areas. With that in mind, this section touches on concepts from the following DARA functional areas:

- Discovery (Functional Area 4)
- Data (Functional Area 2)
- Structural Metadata (Functional Area 3)

4.3.3 STAKEHOLDER PERFORMANCE GUIDE

Table 7. Stakeholder Performance Guide – Applications/Services

STAKEHOLDER PERFORMANCE GUIDE			
SECTION – APPLICATIONS/SERVICES			
Role	Responsibility	Approach	Benefit
Executive Leadership	• Provide leadership support for system integration • Approval of investments in shared services	• Sponsor/Champion the business, policy, and mission needs for interoperability to satisfy the mission requirements.	• Provide guidance, documenting boundaries and mission value to reference architecture. • Documents the mission needs for interoperability and information sharing. • Documents the roadmap/plan to maintain progress toward interoperability goals and to coordinate plans across departments and agencies.
Program Manager	• Define how DARA requirements are incorporated into program plans and delivery schedules	• Incorporate application/services architecture concepts and standards from DARA into program design and implementation plans. • Assist in the development of the resources, scope, and implementation and work plans needed to achieve interoperability goals • Assess the resource, scope, and implementation and work plans needed to achieve interoperability goals. • Communicate the scope, assumptions, and dependencies to Solution Architect.	• Incremental and continuous movement towards the DARA whole of government services. • Documented work plan and roadmap that leads to interoperability. • Coordinate with all stakeholders the desired interoperability architecture artifacts descriptions and scope.
Solution Architect	• Incorporate DARA features, capabilities and standards into solutions architecture	• Incorporate application/service architecture concepts and standards from DARA into solution design and architectures. • Assist in the development of the resource, scope, and implementation and work plans to achieve scope. • Work with counterparts to define implementation details of interoperability.	• Provides input technical integration and resource requirements to achieve scope of adoption of the reference architecture. • Provides technical expertise to the development of application/service architecture and designs.

4.3.4 PARTICIPANT ROLES FOR APPLICATIONS/SERVICES

The Data Aggregation community is comprised of multiple Departments, Agencies, Components, and other organizations, and each plays multiple roles over time and in the context of their mission scenario. For the purposes of this document, we have identified two broad roles to clarify responsibilities:

- **Provider** – This term describes responsibilities relating to services and applications that enable discovery and access to an organization's data that is made available in correlated or raw forms. These organizations are the owners of the mission "master data".
- **Consumer** – This role is for an organization or individual that is searching, discovering or receiving data from another participant in the data aggregation community.

Table 8 summarizes the roles and responsibilities as related to application and services.

Table 8. Participant Roles and Responsibilities (Applications/Services)

DATA AGGREGATION ROLE	CORE APPLICATIONS/SERVICES RESPONSIBILITIES (Mandatory)	TARGET STATE APPLICATIONS/SERVICES RESPONSIBILITIES (Desired)
Provider	• Enable discovery and data interchange services • Enable data enrichment with entity resolution and data correlation	• Expose the DARA standard APIs for discovery and data interchange • Receive back "enriched data" where applicable
Consumer	• Retrieving data within mission parameters and data sharing agreements • Meet the audit requirements specified by the provider for this data	• Consume the DARA standard APIs for discovery and data interchange

4.3.5 APPLICATIONS/SERVICES MATURITY

As the DARA community matures in data sharing capabilities, organizations when acting as providers advance their capabilities to share data in more complex and meaningful ways. The primary maturity model element within the scope of "Applications/Services" is discovery with security considerations as applied by provider organizations. The maturity model change data management described in section 4.1.7 may also be implemented by applications and services such as data interchange services, which are described in Section 4.2.6.

4.3.5.1 BASIC DISCOVERY (LEVELS 1–3 OF APPLICATIONS/SERVICES MATURITY)

The first levels of maturity for DARA applications/services evolve from basic search of a dataset to a configurable federated search using a specific agency-adopted service contract. Programs with these levels of applications/services maturity do not meet the minimum requirements for the definition of a "data aggregation program," which include capabilities to enable automated correlation; however, they may be capable of sharing data with other data aggregation programs and may continue to enable basic data discovery. The basic discovery services build upon supporting elements of the maturity model, such as the access controls, change data management (data interchange) and transport/infrastructure services.

4.3.5.2 DISCOVERY (LEVEL 4–5 OF APPLICATIONS/SERVICES MATURITY)

As DARA applications/services mature, their discovery capabilities evolve to include advanced search of entity-resolved, aggregate data stores with predictive and prescriptive guidance and attribute-highlighted Entity Map results, configurable to federate from systems. These services enable further analysis, exploitation and collaboration among human users, which results in further enrichment of the data. The key difference between levels 4 and 5 of maturity is the use of proprietary or community-adopted services to the use of open service standards and the proposed, future DARA Correlated Entity Data Standard described in Section 4.1.5. The discovery services build upon increasingly mature supporting elements of the maturity model, such as the access control model, change data management (data interchange), transport/infrastructure services, and system performance.

4.3.6 DATA INTERCHANGE SERVICES

Data Interchange services represent data aggregation functionality that enables interagency information sharing, assuming the whole-of-government scope. These may be considered external facing services, and may also be consumed within a given department or agency. The data interchange services described here are the implementation of the change data management capabilities discussed in Section 4.1.7. The data interchange services will rely on supporting transport/infrastructure storage services discussed in Section 4.6.7 for persistent data storage.

4.3.6.1 DATA PUBLISHING/EXPORT SERVICE

The Data Publishing and Export Service enables the sharing of data (entity maps and raw data) within a data aggregation system to be shared with other information sharing partners (and other data aggregation systems). This may include one or more protocols for data interchange, and should ensure that exported data is tagged so that consumers or recipients of published/exported data will be able to continue enforcing the corresponding rules for data access, discoverability, and handling. The Data Publishing and Export Service will function as the external interface to, and consumer of the infrastructure service for persistent data storage.

4.3.6.2 DATA ACCEPTOR/INGEST SERVICE

The Data Acceptor/Ingest service enables ingest of data (either entity maps or raw data) into a data aggregation system. This service should include validation of data tags, where data with invalid tags would be rejected and may include transformation of ingested data in support of internal system requirements for performance or efficiency. The Data Acceptor/Ingest service will function as the external interface to, and consumer of the infrastructure service for persistent data storage. Once data has been validated and ingested into a data aggregation system, it can be made available as appropriate for discovery, retrieval, correlation, and enrichment with data from other sources.

4.3.7 DISCOVERY AND DATA ENRICHMENT SERVICES (FUNCTIONAL AREA 4)

Discovery and Data Enrichment services are the primary capabilities of a data aggregation system. They enable correlation, discovery, and retrieval of information within the aggregated data environment to authorized consumers. Data enrichment can occur as a result of correlation with data from other sources, and as a result of interaction with other analytic processing and collaboration services, which may include feedback and value-added data from human users of a data aggregation system.

4.3.7.1 ENTITY RESOLUTION AND DATA CORRELATION SERVICE

This service supports the process of identifying relationships between entities or determining whether two or more references to real-world objects are referring to the same object within and across disparate data sets. The primary consumer of this service is the Data Acceptor/Ingest Service which attempts to determine if any new or updated data being ingested into a data aggregation system are already identified within an existing entity map. This may include resolution of entities contained within one entity map with another, or correlation of entities between raw datasets. This service should also enable consuming services or human users to "manually" add, update or remove linkages between data. This service will also provide notifications to data providers that there may be two or more entity identifiers assigned to the same entity.

4.3.7.2 CONTENT (ENTITY) EXTRACTION SERVICE

This service supports the extract of named entities (such as names of people, places and things) from unstructured text, and is also known in the field of natural language processing (NLP) as either named entity recognition or information extraction.[11] This service is included even though the emphasis of this initial version of the DARA is on structured data since it is a basic, commonly used service and supports discovery services. Future versions of the DARA will expand on the incorporation of new concepts for unstructured data sharing.

4.3.7.3 ENTITY SEARCH (DISCOVERY) SERVICE

This service supports the ability to discover data and/or metadata about entities within the entity knowledge base (entity map) of a data aggregation system using query selection criteria that includes one or more entity attributes.

4.3.7.4 FULL TEXT SEARCH (DISCOVERY) SERVICE

This service supports the ability to discover data and/or metadata about text content (whether stored in structured or unstructured sources) that satisfies specific query selection criteria.

[11] http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/muc_7_toc.html

4.3.7.5 PERSISTENT SEARCH AND ALERTS (SUBSCRIPTION) SERVICE

The alerts (subscription) service provides the ability for consumers to be notified when information matching a collection of search criteria is added to, or changes within a data aggregation system. Persistent Search provides users with the ability to save search criteria so they can be executed or modified at a later time, either upon demand or as an alert (subscription). The service described here was primarily focused on the user as the consumer; however, it may also support change data management, as described in Section 4.1.7 and Data Interchange in Section 4.2.4.

4.3.7.6 COLLABORATION SERVICES

This service provides the ability for interaction between human users of a data aggregation system to collaborate with each other and preserve the results of their analysis within the data aggregation system. This is also intended to include a capability for enabling human validation of the results of automated entity resolution and data correlation, and adding the results of analysis in the form of comments, notes, or annotations to data within a data aggregation system.

4.4 SECURITY AND PRIVACY (FUNCTIONAL AREAS 3, 5)

4.4.1 FUTURE STATE: SECURITY AND PRIVACY

An optimized approach to Security and Privacy includes implementation of several advanced and emerging concepts including attribute-based access control, automated data tagging that maintains security attributes or dynamically changes them as systems and users enrich data, and a consistent collection of roles and responsibilities alongside a common security model. The result is interoperability of systems and data such that security and privacy models are maintained while appropriate access, with strong consideration to privacy, civil rights, and civil liberties, is seamlessly granted as information is exchanged between organizations and accessed by analysts; interoperability does not occur at the expense of security or privacy, or by invalidating organizational and Federal security or privacy requirements. Ultimately, an optimized data aggregation system will present fully correlated entity maps formatted using open standards with granular, attribute based access controls using “data tags” (resource attributes) that are generated with only limited manual intervention. Fine-grained access control, enabled by Identity and Access Management (IdAM), allows the system owner to manage the system risk by safeguarding use of information without hindering responsible sharing of mission information.

The term “security” is exceptionally broad and means many things to many people. In the context of the DARA, we examine specifically the Identity and Access Management[12] (IdAM) aspect of

[12] IdAM supports, and is supported by, a number of federal directives and issuances, including the CAP Cybersecurity Goals, EO 13587, the National Strategy for Information Sharing and Safeguarding (NSISS), the issuances from the SISS SC, OMB 04-04 and 11-11, plus various CNSS issuances and Intelligence Community directives and standards.

security,[13] which is the most user-facing. IdAM is about how a system interacts with its users. Identity Management is focused on knowing who it is (as well as the individual's basic characteristics) that is interacting with a system or data. Access Management is focused on determining whether or not that individual should be permitted to interact with a specific resource in a specific way. IdAM needs to balance the need to safeguard the system and its data with the need to responsibly share information and enable the mission.

Identity Management and Access Management are two primary Service Types in the Federal Identity Credential and Access Management Roadmap and Implementation Plan ("FICAM Roadmap"). The FICAM Roadmap, Part A, is the Federal Government segment architecture for Identity, Credential, and Access Management (ICAM)[14] and IdAM.

To maintain safeguards and effectively address privacy, civil rights, and civil liberties concerns, partnering departments and agencies must also work with appropriate privacy, civil rights and civil liberties professionals to develop guidance to address the potential impact on individual rights and liberties created by new models of data aggregation and discovery that are not part of the traditional information sharing agreement development process.

Considerations for other dependencies that data aggregation systems that align with other NSISS goals regarding data tagging and FICAM that once the implementation plans are completed, seek to promote attribute-based access data-level tagging based on open standards. Security that incorporates the capabilities prescribed in the FICAM[15] Roadmap and PO#4 Implementation Plan[16] allows system owners to build robust security appropriate to their mission needs that complements other network and cybersecurity activities in a way that meaningfully manages the risk of the system both independently and in conjunction with its network environment. Similarly, activities occurring under PO#2 will help to assure that information sharing and safeguarding actions occur with due consideration and common requirements, including privacy, civil rights, and civil liberties, while still allowing flexibility to meet mission needs.

4.4.2 SECTION CONTENTS

This Security and Privacy section includes components of four functional areas from the Data Aggregation Maturity Matrix as well as other relevant concepts in the following order:

- Stakeholder Performance Guide

[13] Other significant aspects include Certification & Accreditation (now called Assessment & Authorization), FISMA compliance, Physical Security, Network Security, Communications Security, and many others. These aspects are outside of the scope of this document.

[14] http://www.gsa.gov/portal/content/105233?utm_source=OCM&utm_medium=print-radio&utm_term=icam&utm_campaign=shortcuts

[15] Federal Identity, Credential, and Access Management (FICAM) Roadmap and Implementation Guidance, Version 2.0. December 2, 2011. Federal Chief Information Officers Council. The guidance is available at http://www.idmanagement.gov/sites/default/files/documents/FICAM_Roadmap_and_Implementation_Guidance_v2%200_20111202_0.pdf

[16] Such measures should also be reflected in the departmental implementation plan developed by the organization.

- DARA Functional Areas
  - Access Control (Functional Area 4)
  - Structural Metadata (Functional Area 2)
- Other Considerations

Figure 4, below, shows Security and Privacy with respect to other DARA Functional Areas.

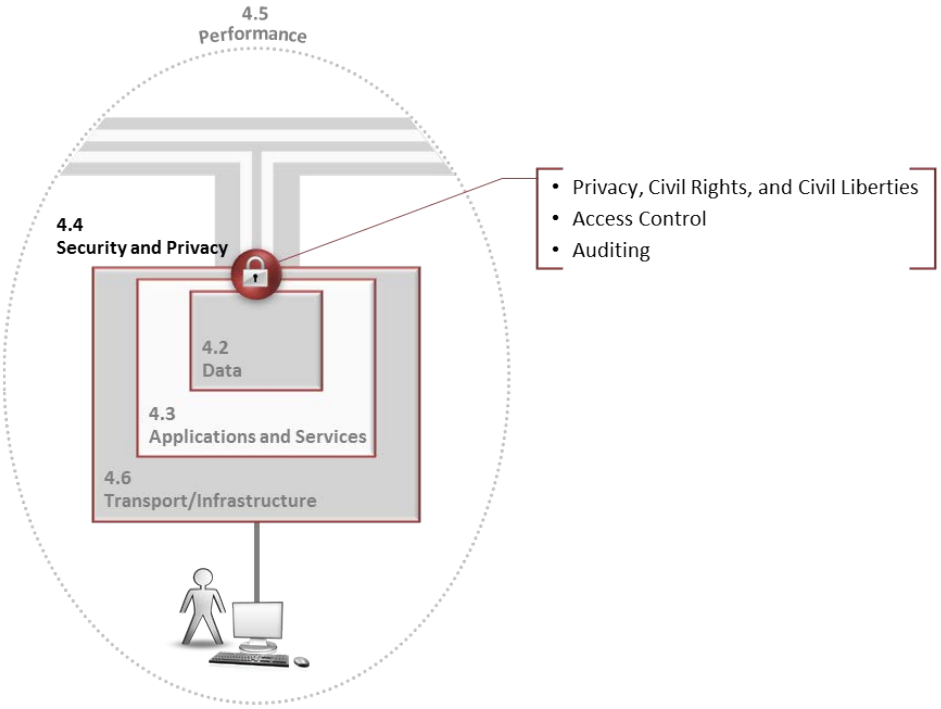


Figure 4. Security and Privacy with respect to other DARA Functional Areas

4.4.3 STAKEHOLDER PERFORMANCE GUIDE

IdAM addresses the policies and technical practices defined by a data owner, vetted by governance and oversight bodies, and enacted by a system owner to protect the information contained in the system. These policies and technical practices must be incorporated into the business practices of the system owner, implemented in the technical capabilities of the system, and enforced as user's access and use the system. As IdAM capabilities become more robust for identifying users and their business purpose in accessing system information, the security model for the system should evolve to take advantage of the additional opportunities for safeguarding system information through fine-grained access control.

Table 9. Stakeholder Performance Guide – Security and Privacy

STAKEHOLDER PERFORMANCE GUIDE			
SECTION – SECURITY AND PRIVACY			
Role	Responsibility	Approach	Benefit
Executive Leadership	• Identify appropriate access, use, and retention policy for system data necessary to ensure responsible information sharing, with consideration to security, privacy, civil rights, and civil liberties requirements, according to mission need.	• Understand the 1) mission need for system information; 2) business processes that incorporate the system information; 3) severity of risk of unauthorized disclosure; and 4) legal and policy restrictions regarding access to and use of the data.	• A clear statement of information sharing policy can be vetted through the relevant stakeholders and then digitally implemented within mission systems to efficiently execute the mission.
	• Ensure risk management function for the organization is established and applies repeatable, consistent evaluation criterion that address security, privacy, civil rights, and civil liberties issues.	• Risk management function should be staffed sufficiently and empowered to reconcile interests of stakeholders. • Clear risk management criteria formed with input from all relevant stakeholders (security, privacy, CR/CL, mission owners).	• Provides consistent feedback that can be incorporated for system design and avoids delays from inability to plan due to ambiguous guidance or interference from unsatisfied stakeholders.
	• Embrace the use of reusable, shared services for IdAM and security capabilities within the agency, and ensure Enterprise Architecture (EA) provides for adoption of federal shared services, particularly IdAM and security services, as they become available.	• Designate organizational Executive Agents responsible for implementing IdAM and Security EA and policy. • EAs represent organization at relevant intergovernmental committees, governance bodies, and WGs.	• Develops acquisition strategy that enables transition of solutions to repeatable shared services.
	• Leverage organizational enterprise architect and direct the inclusion of relevant IdAM and security standards in organizational IT acquisition actions by holding systems accountable for EA compliance.	• EA function should be involved in organizational process for approval of systems to ensure EA for IdAM and Security (services and standards). • Engage organizational acquisitions and procurement functions to ensure contractual commitments and acquisitions are consistent with IdAM and Security EA and implementation plans.	• Ensures that system planning incorporates appropriate guidance from an early stage to avoid delays or wasted expenditures resulting from noncompliant system architecture can be more easily mitigated.

STAKEHOLDER PERFORMANCE GUIDE			
SECTION – SECURITY AND PRIVACY			
Role	Responsibility	Approach	Benefit
Program Manager	• Ensure access policy requirements for security, privacy, civil rights, and civil liberties for the system information are included in system acquisition, tech refresh actions, and system engineering lifecycle.	• Identify access policy rules that have been enumerated for information contained in the system.	• Ensures flexibility and adaptability of systems to incorporate upcoming capabilities.
	• Ensure compliance/evaluation/approval of the system in accordance with the organizational risk management framework.	• Program manager actively engages with relevant governance bodies from system planning phase onward.	• Expedites development by coordinating risk management requirements into system planning and design phase rather than waiting for approval after build is complete.
	• Ensure requirements for relevant IdAM and security standards and EA guidance are included in system acquisition, tech refresh actions, and system engineering lifecycle.	• Give EA organization visibility into each phase of system lifecycle. • EA communicates emerging requirements to program managers.	• Ensures that solutions are engineered or selected to meet all relevant requirements from the planning and design phase. • Ensures that the solution is designed and sufficiently technically implemented to provide flexibility to interoperate with emerging IdAM and Security capabilities without the need for extensive re-engineering.
Solution Architect	• Ensure solution roadmap aligns with FICAM Roadmap and NSISS PO#4 Implementation Plan.	• Detail functionality for currently available capabilities and provide PO&AMs demonstrating alignment for future capabilities.	• Ensures flexibility and adaptability of systems to incorporate upcoming capabilities.
	• Ensure solution meets requirements of organizational risk management framework and security, privacy, civil rights, and civil liberties law and policy.	• Clear system with risk management function during planning stage. If system is operational, coordinate roadmap to satisfy RM function.	• Expedites development by coordinating risk management requirements into system planning and design phase rather than waiting for approval after build is complete.
	• Implement solution that is compliant with EA model for IdAM and Security as well as organizational FICAM implementation plans.	• Solution is described in terms of functional and technical requirements, which are mapped to service types and components of the relevant EA model.	• Ensures that solutions are engineered or selected to meet all relevant requirements from the planning and design phase.
	• Implement solution with sufficient interfaces to take advantage of enterprise IdAM and security services.	• Interfaces are defined sufficiently to show interoperability of system with repeatable shared services and standards.	• Ensures that the solution is designed and implemented to provide sufficient technical flexibility to interoperate with emerging IdAM and Security capabilities without the need for extensive re-engineering.

4.4.4 PARTICIPANT SECURITY AND PRIVACY ROLES

Significant stakeholders in IdAM are diverse and include Data & Mission Owners, Program and Project Managers, System Owners, Enterprise Architects, Information Assurance, and even Procurement personnel.

Table 10 summarizes the roles and responsibilities as related to security and privacy.

Table 10. Participant Roles and Responsibilities (Security and Privacy)

DATA AGGREGATION ROLE	CORE SECURITY AND PRIVACY RESPONSIBILITIES (Mandatory)	TARGET STATE SECURITY AND PRIVACY RESPONSIBILITIES (Desired)
Provider	• Determining, in plain English, the access control policies that are grounded in law and policy for security, privacy, civil rights, and civil liberties for the data they maintain stewardship over. • Determining, in plain English, policies and practices for data use limitation, retention, and redress. • Tag data with NSISS developed standardized access and discovery tag classes which will link to the various technical specification. • Focus on the IdAM service types inherent to their system, and should leverage reusable services where applicable.	• Attribute level access based on open standards for access categories, high flexibility in assigning user credentials, and automated security procedure.
Consumer	• Handle data from various governmental agencies as subject to the original authorities regarding operational security, as well as the privacy, civil rights, and civil liberties of individuals and organizations for which the data pertains. • Respect and enforce data access tags. Providers will likely caveat source data with various access restrictions, and any operations on the source data must appropriately propagate those access restrictions to the resulting entity maps. • Any data that is enriched, any system is subject to adjudication on its data safeguarding strategy . • Structure system calls by passing appropriate user attributes to authenticate requests.	• A correlation system must create new data subject to some hybrid of the source data's controls that does not lessen appropriate use restrictions or policies. The resulting data should be tagged with access controls in an automated fashion, to allow for highly granular access rules (leveraging "cell-level" security features offered by many cloud storage platforms) without a preponderance of manual effort.

4.4.5 PRIVACY, CIVIL RIGHTS AND CIVIL LIBERTIES

Americans are protected in the development and use of the ISE, including the acquisition, access, use, storage, and retention of personally identifiable information. The resulting Guidelines to ensure that the Information Privacy and Other Legal Rights of Americans are protected in the Development and Use of the Information Sharing Environment (ISE Privacy Guidelines)[17] were

[17] The ISE Privacy Guidelines may be found at http://ise.gov/sites/default/files/PrivacyGuidelines20061204.pdf. For additional resources relating to the P/CRCL protection framework for the ISE, please refer to http://ise.gov/privacy-civil-rights-and-civil-liberties-protection-framework.

approved by the President and issued by the Program Manager for the ISE (PM-ISE) on December 4, 2006.

The NSISS has also identified PO #2 that supports the development of guidelines for information sharing and safeguarding agreements to address common requirements, including privacy, civil rights, and civil liberties, while still allowing flexibility to meet mission needs. Under the PO, a Privacy and Information Technology Working Group developed The Framework of Considerations for Streamlining the Information Sharing and Access Agreement Development Process. The DARA working group anticipates that Departments and agencies have developed their own agreement development processes, which have resulted in a variety of agreement development processes but that, over time, the processes will converge based on the work accomplished under PO #2. Without clear and consistent agreements and guidelines, there is a lack of trust between partners regarding implementation of information access, protection, and policy requirements and dispute resolution processes. Trust can be attained by imbedding fair information practice principles into the development of any source system or information sharing initiative. We anticipate this goal to be reached as an iterative progression. Initially, improvements in new and existing bilateral agreements when common requirements are vetted and processes have been adopted. Following the initial adoption, additional requirements for multilateral and federated agreements will be included as the DARA adoption matures. When optimized, the data aggregation architecture, particularly through the implementation of data tagging and increased maturity in access controls and interoperability security models, enforces compliance with privacy, civil rights, and civil liberties requirements. Data tagging, in particular, enables privacy, civil rights, and civil liberties controls by maintaining the information, such as security markings, resource identifiers, lineage, provenance, sensitivity, retention, use limitations, and other handling requirements as organizations share the data.

The Framework of Considerations for Streamlining the Information Sharing and Access Agreement Development Process and Incorporating Privacy, Civil Rights and Civil Liberties Best Practices (Framework) is an effort to define a common procedure for developing information sharing and access agreements (ISAAs). NSISS Priority Objective 2 states: *"Develop guidelines for information sharing and safeguarding agreements to address common requirements, including privacy, civil rights, and civil liberties, while still allowing flexibility to meet mission needs."* In order to streamline the development process and to promote best practices that also conform to the DARA, the Framework will recommend preliminary steps and identify key privacy, civil rights, and civil liberties issues to be considered early in the development of ISAAs to avoid delayed or derailed agreements.

Privacy, civil rights, and civil liberties concerns arise when information is actually shared or when an exchange is implemented. Under DARA, privacy, civil rights, and civil liberties concerns and policies are always valid and applicable and when organizations plan an exchange, privacy, civil rights, and civil liberties professionals must be engaged early in the process. The DARA itself is a reference architecture, does not create new capabilities, but instead leverages existing and possible future capabilities that would evolve with or without DARA implementation, and is not directing that an exchange occur. Therefore, the DARA does not create any new privacy, civil rights, and civil liberties policies that would contradict or supersede existing privacy, civil rights, and civil liberties policies and legislation.

4.4.6 ACCESS CONTROL

The DARA maturity matrix defines enhanced capabilities for implementing access control. Access control occurs at the intersection of policy, legal authorities, data (this section), services, tools and technical standards. The technology has evolved to where certain implementation models are understood, but challenges still remain in policy and procedures. This section deals with the data aspects that are needed to support a community security model, specifically determining access control policy based on data tags, subject attributes, and environmental context. The standards and practices here move the community to a faster engagement across organization boundaries by setting out common concepts and language. Ultimately, access control is a discipline that requires a tight cooperation and coordination between policy makers, data providers and data consumers to define clear access rules and enforce those rules precisely.

The ability to tag data provides a key foundation for access control as it provides the information necessary for an access control system (rules engine) to make an authorization determination based on the metadata for that particular data. The metadata, also known as resource attributes, can be evaluated during the course of a policy decision to determine if a subject (based on their attribute) can be permitted to access said information. In the scope of this document, subjects can be both individual human users and systems (such as devices, applications, processes, etc.) The most basic tagging happens at the data set level, while higher maturity levels apply access tagging to rows or correlated data structures. The highest level of maturity works on the lowest level of data granularity, which varies based on the underlying data storage technology. This also assumes that data tags, as used to enforce access control, are applicable for data-at-rest and data-in-motion. Access Control therefore involves linking data categories and tagged data, identifying user or system attributes that will be evaluated by policy and enforcing that access through a set of policy rules.

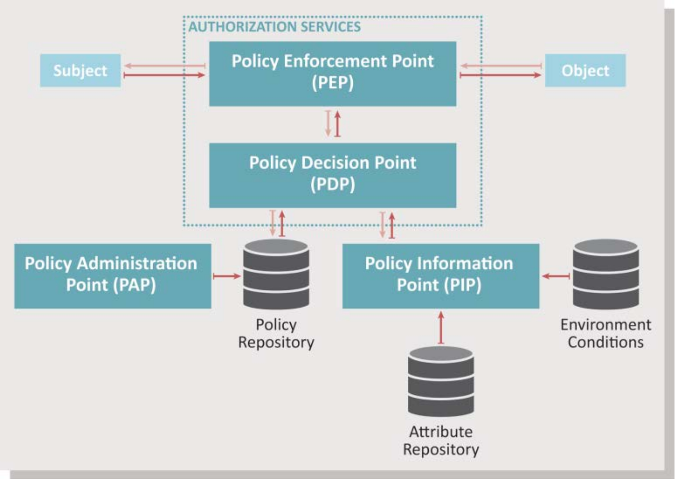

AUTHORIZATION SERVICES
Subject
Policy Enforcement Point (PEP)
Object
Policy Decision Point (PDP)
Policy Administration Point (PAP)
Policy Repository
Policy Information Point (PIP)
Environment Conditions
Attribute Repository

Figure 5. ABAC Functional Architecture (Taken from NIST 800-162)

ABAC is the Intelligence Community (IC) standard and an important part of the NSISS that enables greater interoperability and information sharing. To satisfy the access control use cases for information sharing, the DARA recommends a capability consistent with the *Authorization Services* container, as taken from NIST 800-162[18] of Figure 5. The *Attribute Repository* connected to the *Policy Information Point (PIP)* is completely synonymous with an Attribute repository capability that stores person and system attributes. All person attributes required by the PIP to fulfill an authorization decision will come from the Attribute Hub. In some cases, *Environmental Conditions*, such as risk levels associated to users (perhaps because they have risky—yet required—application entitlements), can be supplied via the Attribute Hub as well, but is expected that most environment conditions will be supplied via runtime from other sources. The Policy Enforcement Point (PEP) should be deployed as closely to the data objects as possible, while the Policy Decision Point (PDP), responsible for evaluating policy based on data object, subject attributes, and environmental conditions, can be deployed in a hybrid centralized and decentralized model, based on the underlying technologies and performance requirements of the data. The PEP is responsible for complying with, and enforcing, all PDP decisions and obligations.

Level 3 is the target for DARA v1.0 participants. While ABAC is not mandatory for maturity level 3, it is desired. The DARA recognizes that not all organizations have the capability to tag data at a level that is granular enough to enable ABAC. Therefore, we describe Level 3, below, in a manner that may include some legacy authorization and access control schemes for those organizations that are still evolving to ABAC[19] in accordance with the NSISS.

Level 1 to Level 2 – Limited or System Level Access – Traditional access controls are implemented at the system or data set level. They provide no granularity within the data collection and as a result, may limit the degree of data sharing due to the coarse controls, or even result in sharing too much. In these levels, from a data aspect, there is limited data tagging for access. Data tagging may be applied at the data set level or may be applied at the record level uniformly across an entire dataset (i.e., very coarse data tagging) with a system-wide "system high" access control requirement. At this maturity level, access is generally authorized system-wide in advance of need through agency security and access authorization procedures.

Level 3 – Role-Based or Map Level Access with Tagging Capability – Controlling access to entities (or "rows" in a traditional database) can be achieved with rules based on key identifiers or field values. Full implementation of this maturity level requires that systems can tag data for access control purposes prior to being shared externally. This level of access control provides an additional degree of sharing granularity, and may enable access according to roles or groups, but is still cumbersome and difficult to enforce across organizational boundaries. While ABAC is not mandatory for maturity level 3, it is desired. The DARA recognizes that not all organizations have

[18] http://csrc.nist.gov/publications/PubsSPs.html

[19] ABAC is mandated by the Steering Committee for SECRET and TS/SCI, and part of FICAM for SBU. All D/As are being held accountable for implementing it on the higher fabrics and it is expected that this will also be true for SBU eventually.

the capability to tag data at a level that is granular enough to enable ABAC. Level 3 looks for access control to be applied to the data at least at the row level, allowing access control systems (rules engine) to make an authorization decision and permit mission users to access the record if authorized. At this maturity level, access is generally authorized well in advance of need and follows agency security and access authorization procedures.

Level 4 – Attribute Based Access Control – This level enhances access granularity by looking at applying access policies to individual fields (or “columns” in traditional database systems). This often supports varying access levels, such as allowing access according to mission role. This approach does not use a standardized means to externalize authorization, but it does drive access control decisions by evaluating attributes. Within this ecosystem, standards-based approaches can be used, as well as legacy or system-based security models that make decisions based on attributes. In addition, acknowledging that legacy systems require account provisioning, this model supports authorization, in advance of need, following business rule evaluation of attributes to determine the appropriate provisioning requirements to create accounts and assign individual permissions.

Level 5 – Standardized Granular Access – Access at the most granular data element level provides the most sophistication and enables broader data sharing capabilities and implementation of the most specific and granular access control policies. Note that access control policies apply at higher levels than the most optimized access control capability, and the differentiating factor at Level 5 is that fine grained access control is enabled and reflected in access control policies at the most granular data level. This level requires a high degree of data tagging, standardized attribute taxonomies, access rule definition and access rule enforcement. Community standards specify the types and values of data tags, and are understood commonly across all participants therefore enabling authorization to access information on demand at the time of need. DARA v2.0 will clarify these standards, potentially applying the concepts and attempt to align the DARA community data standards and architecture with the IC ITE data standards and architecture.

4.4.7 AUDITING

There is a requirement for services that support the recording, examination, and verification of immutable audit logs for accuracy, potentially in line with recommendations made in Report 2014-02 of the Data Privacy and Integrity Advisory Committee on Privacy Recommendations regarding Auditing and Oversight of the DHS Data Framework[20]. The capability to log authorization decisions, accesses, and to trace requests and the movement of data must be present at a minimum. In a future state, requirements could be built for manual, and ultimately, automated analysis of these logs for anomalistic behaviors, activities, or patterns.

[20] http://www.dhs.gov/sites/default/files/publications/dpiac-recommendations-report-2014-02.pdf

4.4.8 OTHER CONSIDERATIONS

4.4.8.1 AUTHENTICATION

Users attempting to access protected resources within data aggregation systems should be subject to, at a minimum, strong authentication using Public Key Infrastructure (PKI). Trusted information sharing partners will be expected to function as identity providers for their own end users. The trusted partner system must perform upfront user authentication and authorization before sending any request to another partner's protected resource on a user's behalf.

Trusted partner systems must support all attributes that are mandatory within each organization which are used for auditing and data disclosure purposes.

Secure authentication may rely on additional infrastructure protection. The NSISS Priority Objective #4 implementation guide and FICAM implementation guidance provides additional details.

4.4.8.2 TAMPER PROTECTION

Tamper protection is extremely important in an information sharing environment. Security procedures will depend heavily upon each organization's enterprise architecture guidance. Organizational and local architecture-compliant mechanisms, such as rejection of data if all required metadata essential to the integrity of the information sharing process is not provided, should be put into place to appropriately guard an agency's own data and any data correlated into the agency's own from another sharing organization. A benefit of correlating data from multiple sources is the lessened likelihood of an adverse impact from bad data or purposeful misinformation. Even as information sharing provides dramatically enhanced mission capability, the tamper protection derived from multiple sources should be preserved.

4.4.8.3 CROSS-DOMAIN TRANSFER

Guards are likely to be placed at various points in the shared information environment, including at each organization to control all incoming and outgoing information, and at the juncture of various security domains, such as the Secret and Top Secret domain.

Guards should leverage granular metadata tags to control information passed through them at a low level. For instance data passing from the TS domain to the Secret domain should have all TS-marked attributes and records redacted, or appropriately obfuscated depending on rules coded by the originated policy decision point.

4.5 PERFORMANCE

4.5.1 FUTURE STATE: PERFORMANCE

Performance, in terms of this reference architecture for data correlation and aggregation, can be broken out into two separate, but related areas: mission enablement, architecture and interoperability implementation. Implementation of the architecture and improvements against an agency's data aggregation system profile should drive increases in an agency's ability to successfully prosecute their mission objectives and should help drive investment in portfolios of capabilities that will further drive this feedback cycle.

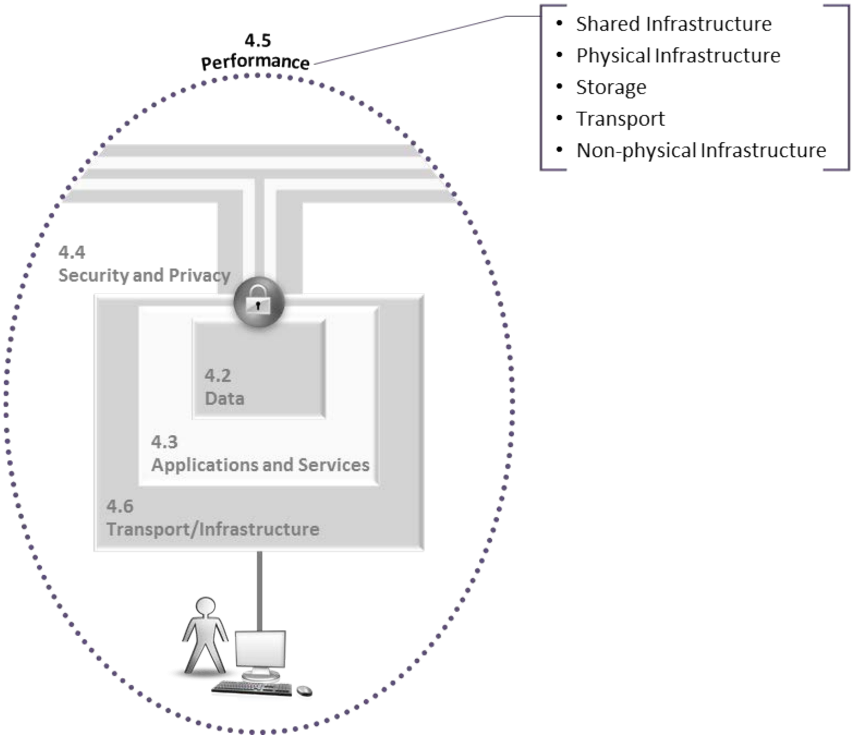
An optimized performance model implies that computing power, storage, and transport capabilities scale at mission speed, systems incorporate new applications and services as they evolve to suit mission purposes, and that data is interoperable and available to consumers when they need it.

4.5.2 SECTION CONTENTS

The Performance section includes two functional areas from the Data Aggregation Maturity Matrix as well as some additional, relevant topics:

- Mission Metrics
- Architecture Implementation Metrics
- Transport/Infrastructure (Functional Area 7)
- Scalability (Functional Area 8)

Figure 6, below, illustrates Performance with respect to other DARA Functional Areas.



- Shared Infrastructure
- Physical Infrastructure
- Storage
- Transport
- Non-physical Infrastructure

Figure 6. Performance with respect to other DARA Functional Areas

4.5.3 STAKEHOLDER PERFORMANCE GUIDE

Table 11. Stakeholder Performance Guide – Performance

STAKEHOLDER PERFORMANCE GUIDE			
SECTION – PERFORMANCE			
Role	**Responsibility**	**Approach**	**Benefit**
Executive Leadership	• Define mission context for investments in portfolio • Identifying funding options for providers to implement IT system changes that only support external consumers.	• Provide overall mission context and expected contribution of programs in oversight portfolio to Program Managers, and align program success to improved performance of business functions. • Establish an Information Sharing IT development funding pool as a means to create prioritized community-wide benefit from initial single agency needs.	• Creates quantifiable measures and expected outcomes (mission and resource impact) of an investment portfolio.
Program Manager	• Define measures of effectiveness and success criteria for projects under oversight • Oversee cost, schedule, and scope of projects in AOR	• Provide clear guidance to Solution Architects for requirements and dependencies of required solutions. • Communicate with executive leadership to foster an understanding of the value of current efforts with the overall mission success.	• Creates clarity as to the value of programs being managed to overall mission effectiveness. • Enables easier management through a better understanding of how measures of effectiveness translate into system requirements.
Solution Architect	• Derive functional and technical requirements given success targets • Oversee technical implementation and provide course corrections as needed	• Analyze program requirements and measures of effectiveness and identify solution elements that will enable the program to meet success criteria. • Create a clear understanding of how the project team is providing value with respect to the overall program and enterprise requirements.	• Demonstrable solution effectiveness, tied directly to executive-level interests which enables an end-to-end picture of how delivered solutions fit into an enterprise-level mission. • Enables clear communication with the project managers and executives regarding schedule and scope of system delivery.

4.5.4 PARTICIPANT ROLES FOR PERFORMANCE

The Data Aggregation community is comprised of multiple Departments, Agencies, Components, and other organizations that each plays multiple roles over time and in the context of their mission scenario. For the purposes of this document, we have identified two broad roles to clarify responsibilities:

- **Provider** – This term is used to describe responsibilities relating to services and applications that enable discovery and access to an organization's data that is made available in correlated or raw forms. These organizations are the owners of the mission "master data".

- **Consumer** – This role is for an organization or individual that is searching, discovering or receiving data from another participant in the data aggregation community.

Table 12 summarizes the roles and responsibilities as related to Performance.

Table 12. Participant Roles and Responsibilities (Performance)

DATA AGGREGATION ROLE	CORE PERFORMANCE RESPONSIBILITIES (Mandatory)	TARGET STATE PERFORMANCE RESPONSIBILITIES (Desired)
Provider	• Sufficiently invest in infrastructure and development activities to preserve or enhance system performance • Maintain configuration management, release notes, and other artifacts that enable community understanding of system changes and performance expectations • Develop MOUs/MOAs as a means to establish provider performance expectations.	• Continuously enhance infrastructure and software, leading to continuous optimized performance • Explicitly add enhanced performance to organization and system goals and objectives
Consumer	• Remain cognizant of community needs when executing large data transfers so that performance is not adversely impacted	• Continually share information on tools, techniques, and practices for enhancing system performance

4.5.5 MISSION METRICS

Driving any implementation of a data correlation or aggregation solution is a set of mission performance requirements that should be directly tied to the success of the system. The data aggregation capabilities described in the use cases outlined in Appendix F are elements of a much larger set of business processes that enable both monitoring and interdiction of known threats as well as the identification of non-obvious links which help with threat discovery. A performance management framework for data aggregation capabilities will include some of the following considerations:

- **Explicit links to intra and inter-agency business processes** – Aggregation and correlation systems built using the interoperability elements found in the DARA will all be used as part of a mission (e.g., involving threat identification or cross-agency intelligence consolidation). By understanding the mission (enabled by correlation and aggregation systems), it is possible to tie the improvement in capabilities directly to an increased ability for those supported system owners to effectively prosecute mission objectives.
  - Metrics dealing with service delivery to mission owners.
  - Improved understanding of data refresh cycle times.
  - Metrics dealing with data movement in correlation and aggregation systems.
  - Scoring and confidence of returned search results.

  - An improved ability to measure the quantity and quality of data made available to partners (not necessarily mission owners).
  - Measuring whether correction or redress notifications are appropriately acted upon.
- **Ties to enterprise architecture investment data** – By tying the DARA-based system implementation to mission outcomes in an enterprise architecture-based investment framework, it becomes possible to roughly quantify the mission value of specific technology initiatives in terms of budgeting and financial data, which improves the ability of system owners to plan for future technology investments.

4.5.6 ARCHITECTURE IMPLEMENTATION METRICS

An early stage in crafting this Data Aggregation Reference Architecture included an assessment of community aggregation and correlation systems, using the Data Aggregation Maturity Matrix, found in Appendix B. Additionally, assessments against the ISA Interoperability Maturity Model (found on the ISE.gov site) can provide an understanding of how investments in a data correlation or aggregation system are contributing to an agency's overall picture of interoperability.

These early stage and current activities provide the basis for measuring benefits from implementing interoperable architecture as data aggregation systems mature. In addition, the Data Aggregation System Profile, in Appendix A, and completed as part of the 5 Step Process described in Section 5, provides a baseline from which future architecture implementation metrics may be measured.

From a system perspective, success at using the DARA can be measured via an improvement over the baseline capability level, either assessed previously or at the time this document is released, of a given program. At a portfolio level, progress in implementing the I^{2}F and DARA can be tracked as a proxy metric until more specific data on improved mission success is available, as the ability to perform and value contributions to the core mission functions due to the interoperability elements of the architecture should be the ultimate measures of effectiveness.

4.5.7 TRANSPORT/INFRASTRUCTURE (MATRIX FUNCTIONAL AREA 7)

As the DARA community matures in measuring and enhancing performance, provider organizations must maintain sufficient infrastructure and connections for data storage, processing, and transport to occur at mission speeds. This requires that provider organizations continually invest in their infrastructure and associated communications capabilities, sometimes provided by third parties, to maintain and enhance performance as data volumes and interoperability activities grow.

Level 1 – Physical or Email Transport. Participants in DARA provide data either by physically transferring it via removable media (sneaker net) or by email. At times, these processes may be necessitated due to security or access limitations. Over time, however, particularly with increased maturity in other functional areas of the maturity matrix enable greater interoperability in security and access control models, physical and email transport for data should be lessened. As this becomes true, organizations invest in infrastructure and transport capabilities so that they do not become inhibitors to improving performance as organizations can share information without resorting to physical or email transport.

Level 2 and 3 – Initial Automated Data Push and Pull. Organizations implement system, and then agency-wide, services that enable automated data push, to consumers, and pull, from providers. The availability of these services and their ability to enhance interoperability increases as DARA implementation leads to greater maturity in Applications and Services, and corresponding functional areas on the maturity matrix. With respect to infrastructure and transport capabilities, organizations make necessary investments in hardware and communications capacity to sustain performance as services and data volumes increase.

Level 4 and 5 – Full Automation and Known Formats. Organizations have implemented agency-wide services that enable full automation of data pushes, to consumers, and data pulls from providers. Organizations continue to enhance interoperability by, first, adopting community-wide standards and formats that enable data interchange on a broad scale. Over time, organizations participate in initiatives to develop open standards and formats, and commit to adopting these standards and formats in their own infrastructure and communications capacity to maintain performance targets.

4.5.8 SCALABILITY (MATRIX FUNCTIONAL AREA 7)

As the DARA community continues adoption and implementation of technologies and methods that enhance interoperability and lead to greater maturity on the Data Aggregation Maturity Matrix, organizations are cognizant and take steps necessary to scale their systems and services in order to continue meeting consumer needs at mission speed. In addition to transport and infrastructure investments described above, organizations proactively identify bottlenecks or areas that slow performance when data volumes or computing requirements spike that prevent them from maintaining consistent levels of performance at all times. At the same time, data consumers continually share tools, techniques, and practices, including technological and non-technical, or procedural, enhancements.

Level 1 – Manual Processes. Organizations have essentially not automated system processes, leading to bottlenecks at any level of usage and an inability to rely on system operations, particularly for external consumers.

Level 2 and 3 – Initial Automation. Organizations have automated some or many processes; however the system's availability is still impacted by unpredictable spikes in data volume or usage. The primary difference between Level 2 and 3 is that, within Level 3 most or all system processes are automated and the data producer could add additional data sources to the system with reasonable assurance that it would handle additional sources of data if the data volume remained constant.

Level 4 and 5 – Full Automation. Organizations have automated all system processes and data producers have confidence that the system will handle additional data sources. The primary differentiator between Level 4 and 5 is that, at Level 4, the system owner cannot accurately specify to what volume of data the system could operate without suffering a performance degradation or availability problem. Systems at Level 5 undergo stress testing in order to document performance parameters at specific levels of scalability measured by data volume or usage.

4.6 TRANSPORT/INFRASTRUCTURE (FUNCTIONAL AREA 6)

4.6.1 FUTURE STATE: TRANSPORT/INFRASTRUCTURE

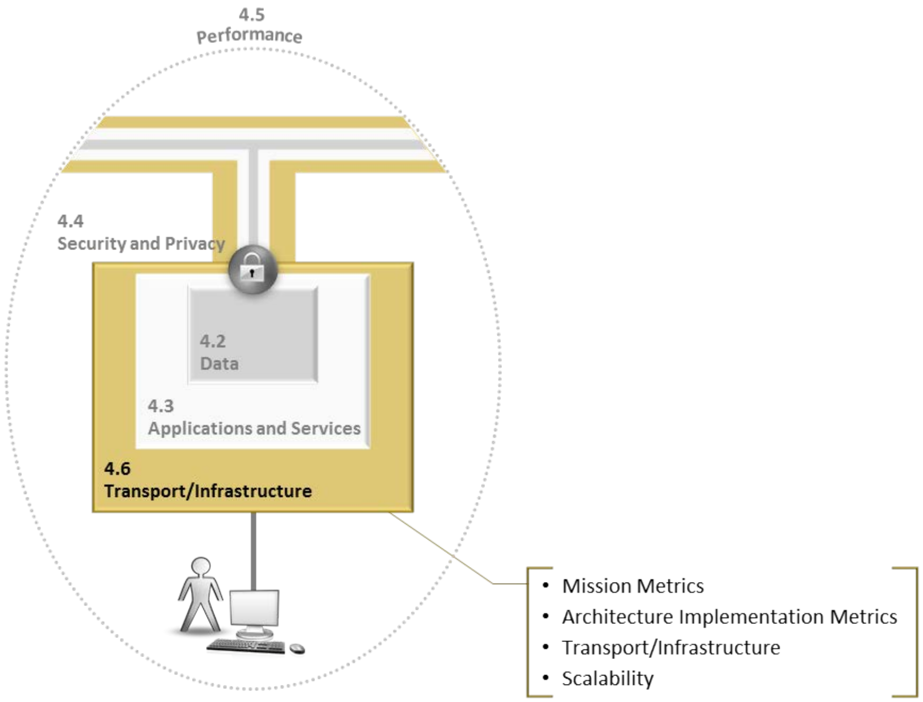
In general, the DARA provides high-level guidance for infrastructure with only the exposed interface document requiring any specificity. Internal infrastructure and methods do not need to be documented unless they affect the interfaces or access to the required data. The infrastructure components that enforce a security and access control model, while outside the scope of this document, must support interoperability as well. An optimized approach to infrastructure includes a well-documented interface that relies on open standards to seamlessly maximize potential interoperability with a variety of systems and organizations.

4.6.2 SECTION CONTENTS

Infrastructure falls directly within the Transport/Infrastructure functional area of the Data Aggregation Maturity Matrix. With that in mind, infrastructure reference architecture is further elaborated in the following topic areas:

- Shared Infrastructure
- Physical Infrastructure
- Storage
- Transport
- Non-Physical Infrastructure

Figure 7, below, illustrates Transport/Infrastructure with respect to other DARA Functional Areas.



- Mission Metrics
- Architecture Implementation Metrics
- Transport/Infrastructure
- Scalability

Figure 7. Transport/Infrastructure with respect to other DARA Functional Areas

4.6.3 STAKEHOLDER PERFORMANCE GUIDE

Table 13. Stakeholder Performance Guide – Transport/Infrastructure

STAKEHOLDER PERFORMANCE GUIDE			
CHAPTER – TRANSPORT/INFRASTRUCTURE			
Role	**Responsibility**	**Approach**	**Benefit**
Executive Leadership	• Define mission context for investments in infrastructure	• Develop justification, rooted in mission and community requirements, for investments in infrastructure • Complete OMB and organizational investment requirements (e.g. Exhibit 53 and 300 requirements)	• Establishes backing and financial support for infrastructure investments that are required to meet mission and community requirements
Program Manager	• Validate infrastructure requirements against mission and organization standards • Oversee implementation	• Provide clear guidance to Solution Architects for requirements and dependencies of required solutions • Communicate with executive leadership to foster an understanding of the value of current efforts with the overall mission success • Document results of infrastructure enhancements	• Creates clarity between investments and mission results • Provides traceability between mission requirements, resources expended, and results • Enables road mapping for future infrastructure requirements

STAKEHOLDER PERFORMANCE GUIDE			
CHAPTER – TRANSPORT/INFRASTRUCTURE			
Role	**Responsibility**	**Approach**	**Benefit**
Solution Architect	• Derive functional and technical requirements given requirements • Oversee technical implementation and provide course corrections as needed	• Analyze program requirements and measures of effectiveness and identify solution elements that will enable the program to meet success criteria • Create a clear understanding of how the project team is providing value with respect to the overall program and enterprise requirements • Design and plan implementation for community standards as they are identified	• Demonstrable solution effectiveness, tied directly to executive-level interests which enables an end-to-end picture of how infrastructure enables interoperability and organizational mission • Enables clear communication with the project managers and executives regarding schedule and scope of system delivery • Maximize interoperability potential, both as provider and a consumer

4.6.4 PARTICIPANT ROLES FOR TRANSPORT/INFRASTRUCTURE

The Data Aggregation community is comprised of multiple Departments, Agencies, Components, and other organizations who each play multiple roles over time and in the context of their mission scenario. For the purposes of this document, we have identified two broad roles to clarify responsibilities:

- **Provider** – This term is used to describe responsibilities relating to services and applications that enable discovery and access to an organization’s data that is made available in correlated or raw forms. These organizations are the owners of the mission “master data”.
- **Consumer** – This role is for an organization or individual that is searching, discovering or receiving data from another participant in the data aggregation community.

Table 14 summarizes the roles and responsibilities as related to transport/infrastructure.

Table 14. Participant Roles and Responsibilities (Transport/Infrastructure)

DATA AGGREGATION ROLE	CORE PERFORMANCE RESPONSIBILITIES (Mandatory)	TARGET STATE PERFORMANCE RESPONSIBILITIES (Desired)
Provider	• Sufficiently invest in infrastructure and development activities to preserve or enhance system performance • Maintain documentation for interface points to share with the DARA community	• Continuously enhance infrastructure to meet evolving mission and interoperability requirements • Adopt community standards as they are identified
Consumer	• Remain cognizant of community needs when executing large data transfers so that performance is not adversely impacted	• Continually share information on tools, techniques, and practices for enhancing system performance

4.6.5 SHARED INFRASTRUCTURE

Each organization although using the same portion of the transport layer may have its own distinct network which it maintains and governs. In order to facilitate data aggregation, organizations will have to make agreements to share one or more resources. These agreements could be instantiated through manual methods or automatically depending on the maturity level of the enterprise. See the Data Aggregation Maturity Matrix (3.2.2.1) for examples of characteristics for each level of maturity. This could be the sharing of physical or nonphysical infrastructure.

4.6.6 PHYSICAL INFRASTRUCTURE

Systems will require some set of physical infrastructure to run on, with appropriately guarded networks. The key elements to data sharing and aggregation within the physical infrastructure are: Networks, Routers, Firewalls, and Server Interfaces (to include access to data centers). Although the reference architecture level should not dictate the details of these elements, at the implementation level each element must be specified in the form of an agency data aggregation system profile in order for future applications/users to access and interface with the Data Aggregation application and services. Internal infrastructure need not be exposed beyond the user/application interfaces.

Architecture artifacts to be developed with regard the elements listed above should include: a network diagram depicting the external interface connections; a list of the applicable technical standards (Technical Standards Profile) of the elements (networks, routers, firewalls, and compute interfaces); and the emerging standards (Technology Forecast) to be considered along with timeframes for future implementations. The use of a profile that is ‘discoverable’ should be considered when documenting standards as this could save costs of development of future systems.

4.6.7 STORAGE

The physical media used to store data for later recall and aggregation. This storage must be able to store and retrieve large amounts of data in a correct, complete, and reliable manner. In addition, storage requirements must be of sufficient size to support long term storage of mission data and its associated metadata as required. The storage may be on the form of a federated environment and should enable the information consumer the ability to access the required data when needed. Use of non-standard or proprietary storage methods/devices should be discouraged. The use of wrappers and/or translator allows data enclaves to use new technologies while still participating in the federated environment.

4.6.8 TRANSPORT

The transport layer provides a medium in which data can be easily accessed, recalled, and used. The layer also consists of three primary domains. It is important to note that shared connectivity will not necessarily mean access between two separate networks. Physical transport constraints should be specified to the point where they are real, and affect the flexibility of the design of instantiated systems. A quick inspection of the Open Systems Interconnection (OSI) model shows that although the Transport layer (Layer 4) involves the reliable delivery of packets between points on a network providing data transfer services to higher layers of the model the lower layer can impact Layer 4. This layer relies on the Network (Layer 3), Data Link (Layer 2) and Physical (Layer 1) layers. How these layers are implemented will affect the performance of the Transport layer and higher layers. Therefore it is necessary to document any considerations (e.g. protocols) and/or constraints at the reference architecture level. Documentation should be included in the network diagrams or standards lists. A link to 'wiki' OSI information is provided here as a resource: OSI Model Information.

4.6.9 OTHER INFRASTRUCTURE CONSIDERATIONS

Other infrastructure considerations includes but is not limited to services and virtualized computing environments including cloud computing environments that organizations use in the process of using and sharing data. The identification of touch points between services and ensuring adherence to recognized standards is also a key aspect of sharing and accessing data when using nonphysical infrastructure. Other considerations to list here include: security service constraints that can restrict the flexibility of instantiated system designs; protocols not called out in the Transport or Physical infrastructure documentation. For security service considerations at the infrastructure level evaluate certification and accreditation requirements early in the process as these can be major impediments to a successful deployment. Specify any other security requirements affecting the infrastructure elements (firewalls, routers, networks) as they will impact the entire DARA architecture and future participants in the architecture.

5 DATA AGGREGATION IMPROVEMENT PROCESS

5.1 HOW TO USE

The Data Aggregation Reference Architecture document is intended to assist in defining the interoperability requirements for data aggregation enterprise investments. For best results, use the **5 Step Approach** to most effectively utilize the DARA, primary architecture frameworks, and other authoritative references throughout the document. Following the 5 Step Approach, completion of the questionnaire for the **Data Aggregation System Profile** in **Appendix A** will assist in identifying the current system's characteristics, techniques, stakeholders and agency partners are currently involved with information exchanges. The resulting profile informs other organizations of the characteristics of the data aggregation system so that they can effectively determine how best to interoperate, provides the basis point from which investment and development decisions, to enhance interoperability, are made, and provides a baseline from which to measure performance.

Use the ISE Information Interoperability Framework Architecture Framework Alignment Grid (Page B-1), Data Aggregation Maturity Model, and Reference Architecture in the following steps.

Table 15. 5 Step Approach

STEP	DESCRIPTION	INTENDED DELIVERABLE/OUTCOME
1	Identify Mission Requirements	Appendix A – Agency Data Aggregation System Profile
2	Perform Maturity Self-Assessment	Appendix B – Self Assessment using Maturity Model
3	Identify the Minimum Requirements for Interoperability	Review and understand the gaps and recommended requirements for progress toward entity correlation and data aggregation goals for your department and agency and the broader ISE community and identify artifacts relevant to interoperability and information sharing.
4	Use the DARA and I2F Framework Grid to Update Applicable Architecture	DARA Appendix B and I2F Appendix B[21] – Update interoperability requirements to address gaps in capability identified in the step 3 to increase the maturity level of functional areas.
5	Build a Plan/Roadmap to Achieve Desired Interoperability Level	Document the set of development efforts, procurement actions or other activities required to make the updates, identified in the previous section, required in the applicable architecture.

5.1.1 STEP 1 – IDENTIFY MISSION REQUIREMENTS

Knowledgeable understanding of organizations' systems: Organizations choosing to interoperate with another organization's system by consuming that system's data can understand important characteristics about the system. This leads to the organization's decision authorities having information about the system that inform actionable decisions with respect to the system's data.

[21] http://www.ise.gov/sites/default/files/FINAL%20-%20ISE_I2F_v0%205.pdf

Their decisions may depend on information about data timeliness, sources, accessibility, metadata, and other aspects Identify mission requirements with specific enterprise reference architecture domain needs, but not limited to:

- Data: Improved data sharing
- New or existing data sharing agreement of what data type (structured, unstructured)
- What classes of entities does this system capability recognize and correlate? (people, time, place, event, organization-centric)
- Structural metadata tagging requirements (metadata tag that supports discovery may also enable flow and access capabilities)
- Improved and/or new application or services required
- Data Interchange services
- Data enrichment, search and discovery services
- Support and infrastructure services
- Improved capability of a system to granularly and interoperable expose subsets of data depending on data's caveats and users' credentials

5.1.2 STEP 2 – PERFORM MATURITY SELF-ASSESSMENT

The Data Aggregation Maturity Matrix, described in Appendix B of the DARA, provides a model by which organizations assess the maturity of their systems using a consistent process that leads to an objective ranking in system maturity for data, structural metadata, discovery, access controls, change data management, transport/infrastructure, and scalability. Organizations' assessment of their systems according to this common maturity matrix enables several activities conducive to participation through DARA and evolution of inter-organization aggregation of data and sharing of correlated entity indexes.

5.1.3 STEP 3 – IDENTIFY THE MINIMUM REQUIREMENTS FOR INTEROPERABILITY

Utilizing sections 4–8, the results of the self-assessment, review and understand the gaps and recommended requirements for progress toward entity correlation and data aggregation goals for your department and agency and the broader ISE community and identify artifacts relevant to interoperability and information sharing.

Section 4.1 – Data: Functionality by the level of correlation, complexity of the resulting records, richness of annotations, and how automated is the data management.

Section 4.2 – Application & Services: As the community matures in data sharing capabilities, organizations when acting as Providers advance their capabilities to share data in more complex and meaningful ways including discovery, change data management, and transport/infrastructure.

Section 4.3 – Security and Privacy: Access control functionality concerns all matters of security and privacy policy within a system and the capability of a system to granularly and interoperable expose subsets of data depending on data's caveats and users' credentials.

Section 4.4 – Performance: Performance for entity correlation and aggregation, can be broken out into two separate, but related areas: mission enablement and architecture, and interoperability implementation.

Section 4.5 – Infrastructure: Transport/infrastructure concerns the interoperability and automation of a system's transport mechanisms.

5.1.4 STEP 4 – USE THE DARA TO UPDATE APPLICABLE ARCHITECTURE

Identify and document the changes, development efforts, or investments that are required to update the data aggregation system's architecture in order to achieve the desired maturity in the Maturity Matrix and enable interoperability. Update interoperability requirements to address gaps in capability identified in the step 3 to increase the maturity level of functional areas.

5.1.5 STEP 5 – BUILD A PLAN/ROADMAP TO ACHIEVE DESIRED INTEROPERABILITY LEVEL

Document the set of development efforts, procurement actions or other activities required to make the updates, identified in the previous section, required in the applicable architecture. Note that OMB may retain these roadmaps/plans in order to maintain progress toward interoperability goals and to coordinate plans across departments and agencies.

5.2 PRIMARY REFERENCES USED THROUGHOUT THIS DOCUMENT ARE:

- *Data Aggregation Reference Architecture* (DARA)
- Appendix B *Architecture Framework Alignment Grid* of the ISE Interoperability Framework[22] (I^2F)
- Accepted Architecture Frameworks (*DoDAF*[23], *GRA*[24], *IC PAG*[25], *TOGAF*[26], etc.)

[22] http://www.ise.gov/sites/default/files/FINAL%20-%20ISE_I2F_v0%205.pdf

[23] http://dodcio.defense.gov/Portals/0/Documents/DODAF/DoDAF_v2-02_web.pdf

[24] https://it.ojp.gov/GRA

[25] https://www.intelink.gov/go/Rd9O0uc

- The *Common Approach to Federal Enterprise Architecture*[27]
- The *National Strategy for Information Sharing and Safeguarding*[28] (NSISS)
- *Federal Identity, Credential, and Access Management* [29] (FICAM) Roadmap and Implementation Guidance
- *Open Systems Interconnection Tool (OSI)*[30]
- Information Sharing Environment public website (www.ise.gov)
- National Intelligence Exchange Model (NIEM) website www.niem.gov
- The ISE Privacy Guidelines[31]
- P/CRCL protection framework for the ISE[32]

5.3 MATURITY MATRIX SELF-ASSESSMENT RESULTS AND DOCUMENTATION

Organizations' assessment of their systems according to this common maturity matrix enables several activities conducive to participation through DARA and evolution of inter-organization aggregation of data and sharing of correlated entity indexes. These activities include:

- Procurement and technology road mapping and lifecycle planning: Organizations may evaluate themselves at lower levels (e.g., 1) Ad Hoc, 2) Repeatable) on the maturity matrix. Lower ratings do not imply a barrier to participation via DARA, but instead are informative to other participants and enable organizations to manage expectations for using data from organizations and systems that have self-evaluated at a lower maturity. When organizations evaluate a system at a lower level of maturity, the matrix provides guidance for organizations to understand, and plan, the activities or procurement actions that are necessary to move the system to a higher level of maturity. For example, an organization may rate their system a 1 or 2 for scalability may begin procurement planning to increase infrastructure or move to a cloud environment while planning technical activities required to enable an application to operate in a cloud environment and lead to a state that includes "Fully automated support for any conceivable data/usage volume and additional resources."
- Knowledgeable understanding of other organizations' systems: Organizations choosing to interoperate with another organization's system by consuming that system's data can

[26] http://pubs.opengroup.org/architecture/togaf9-doc/arch/

[27] http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/common_approach_to_federal_ea.pdf

[28] http://www.whitehouse.gov/sites/default/files/docs/2012sharingstrategy_1.pdf

[29] http://www.idmanagement.gov/sites/default/files/documents/FICAM_Roadmap_and_Implementation_Guidance_v2%200_20111202_0.pdf

[30] http://en.wikipedia.org/wiki/OSI_model

[31] http://ise.gov/sites/default/files/PrivacyGuidelines20061204.pdf

[32] http://ise.gov/privacy-civil-rights-and-civil-liberties-protection-framework

understand important characteristics about the system. This leads to the organization's decision authorities having information about the system that inform actionable decisions with respect to the system's data. Their decisions may be informed by information about data timeliness, sources, accessibility, metadata, and other aspects. For example, an understanding that a system's data is only current to a certain point may lead to a decision authority requiring some additional verification prior to making an operational decision regarding that system's data. This does not imply that the data is not valuable for discovery, but is purely informational to decision authorities as they decide next actions.

5.4 SUMMARY

The future state envisioned with full DARA implementation is the appropriate availability to participating organizations of raw and correlated data at the speed necessary to identify and counter rapidly evolving threats. This will occur via the broad or complete adoption of community-wide standards as organizations implement the DARA, and the DARA and community systems continue to evolve over the next several years. As the maturity of the data aggregation systems continues to improve through initiatives of individual departments and agencies, and utilizing frameworks like the DARA, the ability to prescribe community adopted interoperability standards and techniques will be realized. The vast knowledge and experience currently in the inter-agency, and lessons learned will assist in development of future versions of the DARA.

If you have questions about the DARA version 1, its use or to provide feedback and lessons learned to the DARA development team, please go to **www.ise.gov** and select the "Contact Us" page.

This page intentionally blank.

APPENDICES

This page intentionally blank.

A. AGENCY DATA AGGREGATION SYSTEM PROFILE

Please provide the following details:

Table A-1 Data Aggregation System Profile

A	GENERAL QUESTIONS	RESPONSES
1	Name of data aggregation program/system:	
2	Sponsoring organization:	
3	Points of contact for follow-up questions:	
	a. Primary POC:	(name, e-mail, phone number)
	b. Alternate POC:	(name, e-mail, phone number)
4	Description and mission use of system:	
B	**DATA**	**RESPONSES**
5	What type of data does this system contain?	(Structured, Unstructured, or Both)
6	What classes of entities does this capability recognize and correlate?	
	a. People-centric:	(yes/no)
	b. Time-centric:	(yes/no)
	c. Place-centric:	(yes/no)
	d. Event-centric:	(yes/no)
	e. Organization-centric:	(yes/no)
	f. Other (please describe):	(yes/no)
7	Is the data Title 50 or non-Title 50 data?	(Title 50/Non-Title 50)
8	Type of "INTs" included in system (HUMINT, SIGINT, IMINT, FISINT, FININT, GEOINT, MASINT …)	
9	Does the system include data on:	
	a. U.S. persons?	(yes/no)
	b. Other special protected classes of individuals?	(yes/no)
	c. Law enforcement data?	(yes/no)
	d. Personally Identifiable Information?	(yes/no)
10	Highest classification of data included:	(Unclassified, SBU, Secret, Top Secret)
11	From which agencies do you receive data feeds? Please list the agencies and the data sources.	(list those where agreements exist and those planned)
C	**TECHNICAL**	**RESPONSES**
12	What is the basic architecture and retrieval methodology used in this system?	(e.g., distributed data with federated search; data ingest with centralized query; hybrid)
13	What is the approximate amount of data, whether distributed or ingested, in the system?	(size estimate; in gigabytes and/or number of records and average attribute size of records)

14	If ingest processes are employed in the system, what is time period of refresh for ingested data?	(Real or near-real time; hourly, daily, weekly, monthly, etc.)
15	Does this program or system ingest data from the following entities?	
	a. Intelligence community?	(consume/produce/manipulate data)
	b. State and local governments?	(consume/produce/manipulate data)
	c. Tribal partners?	(consume/produce/manipulate data)
	d. Private sector entities?	(consume/produce/manipulate data)
	e. International government allies?	(consume/produce/manipulate data)
	f. Other Federal agencies (OGAs)	(consume/produce/manipulate data)
D	**PARTNERS/USERS**	**RESPONSES**
16	Who receives output from the program or system? Examples: • USG senior leaders • IC elements • Non-IC Federal partners • Private sector partners • Foreign government partners • SLT partners	
17	Is this program or system accessible by other partners: • Other Federal D/As? • Foreign government entities? • Private sector entity? • SLT entity?	(agency name)

ADDITIONAL DETAILS

Please provide any additional information (graphics, tables, documentation, or links) of program artifacts that may be helpful to understand scope and purpose of the system.

B.MATURITY SELF-ASSESSMENT

The maturity model is expressed in terms of seven (8) functional areas for data correlation: Business, Data, Structural Metadata, Discovery, Access Control, Change Data Management, Transport/Infrastructure, and Scalability with characteristics established for each level of interoperability (ad hoc, repeatable, enhanced, managed, and optimized) for each interoperability requirement.

For each functional area, determine the maturity level of your mission architecture by moving across each row and matching your current state. During this step you should also note the characteristics of each requirement where the requirement/element maturity is less than your desired level (Ex., your interoperability level is at ‘repeatable’; you need to be at ‘managed’). Note that mission-specific architectures will have different goals for each element maturity level based on the operational needs or organizational policy of the mission architecture.

Instructions: Please mark directly on this matrix for a quick assessment of your system taking into consideration the **themes-based composition** to assist in the assessment. If you feel your system’s current state reflects pieces of multiple categories, mark applicable segments. Then, continue to the Detailed Row Assessment section.

MATURITY LEVELS

① **Ad Hoc:** Initial (chaotic, ad hoc) – the starting point for use of a new or undocumented repeat capability

② **Repeatable** – Documented sufficiently such that repeating the same steps may be attempted

③ **Enhanced** – Defined/confirmed as a standard business process, and decomposed to levels 1 and 2 (the last being Work Instructions)

④ **Managed** – Quantitatively managed in accordance with agreed-upon metrics

⑤ **Optimized** – Management includes deliberate capability optimization and improvement

This page intentionally blank.

Table B-1. Data Aggregation Maturity Matrix

#	FUNCTIONAL AREA	① AD HOC	② REPEATABLE	③ ENHANCED	④ MANAGED	⑤ OPTIMIZED
1	**Business**	Little or no business process definition, including documentation or modeling with no relationships between system requirements and organizational business functions	Business processes, including information flow, are defined with documentation within the organization, enabling simple modeling, with consistent configuration management principles applied, and providing a foundation for system requirements	Business processes, including information flow, are defined with documentation that illustrates relationships to and dependencies on enterprise processes; formalized business process definitions can be understood by external partners leading to understanding of the impact of system requirements changes on the organization, or vice versa	Business processes, including information flow, are defined using open standards (e.g., UML) that enable inter-agency and community interoperability, and are understood by external partners with the ability to model the mission impact, throughout the community, from changes to system requirements	Business processes, including information flow, are defined using open standards (e.g., UML) that enable interoperability across the whole of government with documented understanding of information providers, consumers, and associated relationships and dependencies with documented understanding of system requirements as they affect information providers, consumers, and associated relationships and dependencies
2	**Data**	Raw data with little or no sourcing	Entity records with system-level data tags generated with some manual intervention	Enriched records with record/key-level data tags generated with some manual intervention	Partially correlated entity maps with data tags generated with little manual intervention	Fully correlated maps with granular data tags generated automatically with only manual approval
3	**Structural Metadata**	Basic data structures with inconsistent formatting	Consistent XML-based format with some automated structuring and manual record-level tagging	Consistent agency-adopted format with mostly automated structuring and manual record-level tagging	Semi-automatically tagged at the attribute-level(e.g., automatic tagging with manual approval) with community-adopted metadata format	Data tagged at the attribute-level with open metadata standards
4	**Discovery**	Basic dataset-wide search with ranked results	Basic system-wide search with ranked results	Basic search with ranked results, configurable to federate from any system using a specific agency-adopted service contract	Advanced search with some predictive guidance and Entity Map results, configurable to federate from any system using a community-adopted service contract	Advanced search with predictive and prescriptive guidance and attribute-highlighted Entity Map results, configurable to federate from systems using an open standard

#	FUNCTIONAL AREA	① AD HOC	② REPEATABLE	③ ENHANCED	④ MANAGED	⑤ OPTIMIZED
5	**Access Controls**	None/Physical access controls only, or system-wide ("system high") access control with access provisioned in advance of need.	Collection level (communities, databases, or other system-defined groups of data) access control based on system-specific access categories (Roles, Groups, etc) with access provisioned in advance of need.	Record/Map-level access based on agency-wide access categories with access being provisioned in advanced of need.	Record/Map-level (or better) attribute-based access control based on system-specific attributes with access being provisioned in advance of need.	Record/Map-level (or better) attribute-based access based on community-wide attributes with access being provisioned on demand at the time of need.
6	**Change Data Management**	Manual replacement of all system data with no timestamps	Automated replacement of system data with simple timestamps	Automated replacement of changed record sets with history	Automated, event-driven replacement of changed records with history	Automated, event-driven, sub second replacement of changed attributes with history
7	**Transport/ Infrastructure**	Physical/email transport	System-specific service with mostly automated pushes and pulls	Agency-wide service with entirely automated pushes and pulls	Configurable to operate with any system using a community-adopted proprietary format with entirely automated pushes and pulls	Configurable to operate with any system using an open standard with entirely automated pushes and pulls
8	**Scalability**	The system's (largely manual) processes bottleneck even under normal volume	The system has some manual processes, and has limited support for additional volume of data, data sources, or users	The system is reasonably automated, but spikes in volumes and usage require extra monitoring; could add additional data sources if volume is mitigated	The system is entirely automated, but there is reason to doubt the processes in place could handle a large spike in data/usage, or many new sources	Fully automated support for any conceivable data/usage volume and additional sources

THEMES-BASED COMPOSITION

Discussing the Data Aggregation Maturity Matrix content presents several challenges. Due to a number of repetitive technical themes resurfacing periodically while considering different rows, adequately delineating the focus of each row becomes difficult. Across rows, relevant details may be added and dropped off as maturity increases, when consistency in considering those details for each stage would be preferable.

In the interest of framing the conversation, the following maturity themes can be considered for each row, both for framing the nature of the row and ensuring thorough consideration of phases a system passes through as it matures.

Table B-2 Themes-Based Composition

THEME	LOW	LOW-MED	MED	MED-HIGH	HIGH
Granularity	System-wide	Domain-wide	Ranked record list	Record/map-level	Attribute/cell-level
Correlation	Raw data	Entity records	Enriched records	Partially correlated Entity Maps	Fully correlated Entity Maps
Automation	Manual/Physical	Some manual	Manual initiation	Human auditing only	Fully automated
Latency	Weeks	Days	Hours	Minutes	Sub-second
Trust	Indeterminate/ Inconsistent	Some minimum ensured	Noted authority for verification	Various descriptors	Fully auditable history
Interoperability	Little or none	System-wide	Agency-wide	Adopted proprietary standard	Open standard

This page intentionally blank.

DETAILED ROW ASSESSMENTS

BUSINESS

We define Business Domain maturity according to the completeness of business process documentation, the ability to model business processes, use of open standards to document business processes, and the understanding and documentation of relationships between organizational, enterprise, community, and whole of government business processes.

① AD HOC	② REPEATABLE	③ ENHANCED	④ MANAGED	⑤ OPTIMIZED
Little or no business process definition, including documentation or modeling with no relationships between system requirements and organizational business	Business processes, including information flow, are defined with documentation within the organization, enabling simple modeling, with consistent configuration management principles applied, and providing a foundation for system requirements	Business processes, including information flow, are defined with documentation that illustrates relationships to and dependencies on enterprise processes; formalized business process definitions can be understood by external partners leading to understanding of the impact of system requirements changes on the organization, or vice versa	Business processes, including information flow, are defined using open standards (e.g., UML) that enable inter-agency and community interoperability, and are understood by external partners with the ability to model the mission impact, throughout the community, from changes to system requirements	Business processes, including information flow, are defined using open standards (e.g., UML) that enable interoperability across the whole of government with documented understanding of information providers, consumers, and associated relationships and dependencies with documented along with understanding of system requirements as they affect information providers, consumers, and associated relationships and dependencies

BUSINESS DOMAIN DEFINED BY THEMES

- **Documentation** – to what extent are business processes defined and documented?
- **Modeling** – to what extent does business process documentation enable modeling, particularly of information sharing and interoperability, allowing estimates on the impact of gaining or losing data sources?
- **Relationships** – how well understood are business process relationships within an organization, between an organization and its larger enterprise, with the community, and with the rest of government?
- **Standards** – did the organization document business processes using standards, and are the standards proprietary or open?

- This row comingles system maturity and organizational maturity as, in this case, system maturity will depend on organizational maturity, for example, system requirements are derived from, among other things, organization business processes following a reference model to document business functions as described in the Federal Enterprise Architecture.
- Documentation is the major factor for this row, particularly the extent to which it is defined and represents relationships and dependencies outside of the organization.
- Configuration management is a critical component to maintaining business process documentation, and should also be considered when assessing maturity.

CURRENT CAPABILITIES

Please rank the system on the above maturity scale with consideration to organizational maturity. If your system does not cleanly fall into any one category, please select portions of the above to briefly describe the state of your organization’s business process documentation and the extent to which system requirements are derived from them. In either case, please provide detail to support your assessment.

LESSONS LEARNED

Please describe, at a high level, some challenges faced in developing your system to this stage of maturity in this functional area.

FUTURE INITIATIVES

If there are any current or planned system developments or organizational documentation efforts that would lead to the system achieving higher maturity levels, please describe. Include timelines or milestones as available.

DATA

We define Data functionality by the level of correlation, complexity of the resulting records, richness of annotations, and how automated the data management. We define correlation as calculated, entity-based consolidation of records.

① AD HOC	② REPEATABLE	③ ENHANCED	④ MANAGED	⑤ OPTIMIZED
Raw data with little or no sourcing	Entity records with system-level data tags generated with some manual intervention	Enriched records with record/key-level data tags generated with some manual intervention	Partially correlated entity maps with data tags generated with little manual intervention	Fully correlated maps with granular data tags generated automatically but still requiring manual approval

DATA DEFINED BY THEMES

- **Correlation** – How correlated is the data?
- **Trust** – What does the user know about the data's source?
- **Granularity** – How fine-grained are the processing notes?
- **Automation** – How much manual tagging is required for correlation to occur?
- Correlation is the major factor for this row.
- This row does not apply to Raw data stores.
- As with change management, proper sourcing details should be implemented early in maturation.

CURRENT CAPABILITIES

Please rank the system on the above maturity scale. If your system does not cleanly fall into any one category, please select portions of the above to briefly describe the state of your system's data. In either case, please provide detail to support your assessment.

LESSONS LEARNED

Please describe, at a high level, some challenges faced in developing your system to this stage of maturity in this functional area.

FUTURE INITIATIVES

If there are any current or planned development efforts on the system towards higher maturity levels, please describe. Include timelines or milestones as available.

STRUCTURAL METADATA

Structural Metadata considers the flexibility of data's structure. In assessing this functional area, we consider the interoperability of the data formatting standards used and the level of automation with which those standards are implemented and enforced. While metadata functionality is leveraged by many other functional areas (access controls, for instance), the goal of development in the Structural Metadata area is to provide the data scaffolding necessary to support the other functional areas that may leverage it. In making this assessment, please avoid considering factors that depend on the richness of the metadata, but are apart from the structural components of the metadata.

① AD HOC	② REPEATABLE	③ ENHANCED	④ MANAGED	⑤ OPTIMIZED
Basic data structures with inconsistent formatting	Consistent XML-based format with some automated structuring and manual record-level tagging	Consistent agency-adopted format with mostly automated structuring and manual record-level tagging	Semi-automatically tagged at the attribute-level with community-adopted metadata format	Data tagged at the attribute-level with open metadata standards

STRUCTURAL METADATA DEFINED BY THEMES

- **Interoperability** – How standardized is the structural format?
- **Automation** – How much manual work is required to format the data?
- **Granularity** – At what level can data be "tagged"?
- Interoperability is the key factor.
- Granularity should be high starting early in the maturation process.
- Automation should mostly follow hand-in-hand with interoperability. Automation may speed up ahead as maturity progresses.

CURRENT CAPABILITIES

Please rank the system on the above maturity scale. If your system does not cleanly fall into any one category, please select portions of the above to briefly describe the state of your system's data. In either case, please provide detail to support your assessment.

LESSONS LEARNED

Please describe, at a high level, some challenges faced in developing your system to this stage of maturity in this functional area.

FUTURE INITIATIVES

If there are any current or planned development efforts on the system towards higher maturity levels, please describe. Include timelines or milestones as available.

DISCOVERY

Discovery functionality concerns the search capabilities of a system, including the granularity of search parameters and search results, the breadth of data sources a system is able to search, and any automated features for predictive/prescriptive search available for discovery in the system.

① AD HOC	② REPEATABLE	③ ENHANCED	④ MANAGED	⑤ OPTIMIZED
Basic dataset-wide search with ranked results	Basic system-wide search with ranked results	Basic search with ranked results, configurable to federate from any system using a specific agency-adopted service contract	Advanced search with some predictive guidance and Entity Map results, configurable to federate from any system using a community-adopted service contract	Advanced search with predictive and prescriptive guidance and attribute-highlighted Entity Map results, configurable to federate from systems using an open standard

DISCOVERY DEFINED BY THEMES

- **Granularity$_P$** – How fine-grained are search parameters?
- **Granularity$_R$** – How fine-grained are search results?
- **Interoperability** – How wide is the net cast by a search?
- **Automation** – How much predictive guidance in search does the system provide?
- Granularity in results will likely stay lower than granularity in parameters; neither will start low, and results won't end high, due to minimal requirements of search functionality and a need for a broad net on search results.
- Granularity for discovery purposes is likely comprehensive: Any system that can handle low granularity for either parameters or results can likely handle all higher levels for parameters or results, respectively.
- Types and levels of predictive guidance will vary from mission need to mission need; wording of the final table should avoid being overly specific. Automation does not need to be high for Discovery.
- Because data aggregation is a priority, interoperability should be prioritized over granularity.

CURRENT CAPABILITIES

Please rank the system on the above maturity scale. If your system does not cleanly fall into any one category, please select portions of the above to briefly describe the state of your system's data. In either case, please provide detail to support your assessment.

LESSONS LEARNED

Please describe, at a high level, some challenges faced in developing your system to this stage of maturity in this functional area.

FUTURE INITIATIVES

If there are any current or planned development efforts on the system towards higher maturity levels, please describe. Include timelines or milestones as available.

ACCESS CONTROLS

Access control functionality concerns all matters of security policy, which should also reflect privacy, civil rights, and civil liberties concerns, within a system. This functional area concerns the capability of a system to granularly and interoperable expose subsets of data depending on data's access rules, data tags and users' attributes. In the table's wording, both data tags (or caveats) and user attributes qualify as "access rules"; in assessing the maturity of the system, only the granularity and flexibility with which access rules can be assigned to data should be considered until the Optimized column. To qualify as Optimized, a system should be equally flexible with the degree to which access categories can be assigned to individuals.

① AD HOC	② REPEATABLE	③ ENHANCED	④ MANAGED	⑤ OPTIMIZED
None/Physical access controls only, or system-wide ("system high") access control with access provisioned in advance of need.	Collection level (communities, databases, or other system-defined groups of data) access control based on system-specific access categories (Roles, Groups, etc) with access provisioned in advance of need.	Record/Map-level access based on agency-wide access categories with access being provisioned in advanced of need.	Record/Map-level (or better) attribute-based access control based on system-specific attributes with access being provisioned in advance of need.	Record/Map-level (or better) attribute-based access based on community-wide attributes with access being provisioned on demand at the time of need.

ACCESS CONTROL DEFINED BY THEMES

- **Granularity** – How fine-grain are access controls?
- **Interoperability$_D$** – How standardized are the access control caveats on the data?
- **Interoperability$_P$** – How standardized are the access control credentials on the person?
- **Automation** – How much manual intervention is necessary to associate security levels? Are they automatically inherited and assumed (erring on the side of over restriction) when possible?

CURRENT CAPABILITIES

Please rank the system on the above maturity scale. If your system does not cleanly fall into any one category, please select portions of the above to briefly describe the state of your system's data. In either case, please provide detail to support your assessment.

LESSONS LEARNED

Please describe, at a high level, some challenges faced in developing your system to this stage of maturity in this functional area.

FUTURE INITIATIVES

If there are any current or planned development efforts on the system towards higher maturity levels, please describe. Include timelines or milestones as available.

PRIVACY, CIVIL RIGHTS, AND CIVIL LIBERTIES

Privacy, civil rights, and civil liberties are critical to ensuring the long-term success of data aggregation and correlation system in a free and open society. These controls are even more important as Departments and Agencies bring together data from across multiple missions.

① AD HOC	② REPEATABLE	③ ENHANCED	④ MANAGED	⑤ OPTIMIZED
No privacy, civil rights or civil liberties tags or policies incorporated into sharing or documented in an MOA or MOU. System of Records Notice or Privacy Impact Assessment do not address sharing. No or very limited compliance with NIST Special Publication 800-53 Appendix J Privacy Controls.	System of Records Notice or Privacy Impact Assessment address sharing in terms of minimal compliance. MOAs or MOUs document access, use, and retention policies. Limited compliance with NIST Special Publication 800-53 Appendix J Privacy Controls.	MOAs and MOUs provide detailed privacy, civil rights, and civil liberties protections. Technical controls provide basic privacy, civil rights, and civil liberties. Audit logs are collected but not analyzed. Moderate compliance with NIST Special Publication 800-53 Appendix J Privacy Controls.	Privacy Impact Assessment provides robust transparency into activities. Granular access controls and data tags enforce privacy, civil rights, and civil liberties policies. Audit logs are collected and analyzed for security concerns. Full compliance of - NIST Special Publication 800-53 Appendix J Privacy Controls.	Multi-front transparency initiative. Audit logs are collected and analyzed for policy compliance. Independent periodic program reviews to determine privacy, civil rights, and civil liberties compliance.

ACCESS CONTROL DEFINED BY THEMES

- **Legal Compliance** – Do the System of Records Notice or Privacy Impact Assessment permit the sharing, access, use, and retention of the data?
- **Technical Controls** – How are the privacy, civil rights, and civil liberties protections and controls documented and enforced?
- **Transparency** – How transparent is the program in its public-facing documentation?
- **Automation** – Are refresh and redress automated?
- **Accountability** – Are audit logs reviewed for policy compliance?

CURRENT CAPABILITIES

Please rank the system on the above maturity scale. If your system does not cleanly fall into any one category, please select portions of the above to briefly describe the state of your system's data. In either case, please provide detail to support your assessment.

LESSONS LEARNED

Please describe, at a high level, some challenges faced in developing your system to this stage of maturity in this functional area.

FUTURE INITIATIVES

If there are any current or planned development efforts on the system towards higher maturity levels, please describe. Include timelines or milestones as available.

CHANGE DATA MANAGEMENT

Change Data Management concerns the ability of a system to receive and/or propagate data updates and deletions within the system of record from which they originate.

① AD HOC	② REPEATABLE	③ ENHANCED	④ MANAGED	⑤ OPTIMIZED
Manual replacement of all system data with no timestamps	Automated replacement of system data with simple timestamps	Automated replacement of changed record sets with history	Automated, event-driven replacement of changed records with history	Automated, event-driven, sub second replacement of changed attributes with history

CHANGE DATA MANAGEMENT DEFINED BY THEMES

- **Granularity** – What data units update independently?
- **Latency** – How long may it take to reflect a change in source data
- **Trust** – What does the user know about the age/source of the data?
- **Automation** – How much manual intervention is required for an update to occur?
- Granularity improves speed. The smaller the update piece, the faster the update piece is to process.
- Automation improves granularity. Because small-scale updates require combing over many pieces of data to assess changes, the process requires higher levels of automation.
- Detailed information on the trustworthiness of data (Trust) is important in the absence of sub-second updates, and reporting on the age/source of data should be easier to implement than fine-grain update capability. Therefore Trust escalation should occur early in the maturation process for Change Data Management.

- Low-latency data transference is difficult on a number of levels, and not always necessary. As such, the maturity of a system will not be predicated on latency for levels 1-4. In order to be truly optimized (level 5), however, sub-second latency is important.

CURRENT CAPABILITIES

Please rank the system on the above maturity scale. If your system does not cleanly fall into any one category, please select portions of the above to briefly describe the state of your system's data. In either case, please provide detail to support your assessment.

LESSONS LEARNED

Please describe, at a high level, some challenges faced in developing your system to this stage of maturity in this functional area.

FUTURE INITIATIVES

If there are any current or planned development efforts on the system towards higher maturity levels, please describe. Include timelines or milestones as available.

TRANSPORT/INFRASTRUCTURE

Transport/infrastructure concerns the interoperability and automation of a system's transport mechanisms. To a point, it also includes the physical connectivity of a system.

① AD HOC	② REPEATABLE	③ ENHANCED	④ MANAGED	⑤ OPTIMIZED
Physical/email transport	System-specific service with mostly automated pushes and pulls	Agency-wide service with entirely automated pushes and pulls	Configurable to operate with any system using a community-adopted proprietary format with entirely automated pushes and pulls	Configurable to operate with any system using an open standard with entirely automated pushes and pulls

TRANSPORT/INFRASTRUCTURE DEFINED BY THEMES

- **Interoperability** – Does the system present data in a standardized way?
- **Automation** – Are pushes and pulls supported?
- Automation should accelerate ahead of interoperability as maturity progresses.

CURRENT CAPABILITIES

Please rank the system on the above maturity scale. If your system does not cleanly fall into any one category, please select portions of the above to briefly describe the state of your system's data. In either case, please provide detail to support your assessment.

LESSONS LEARNED

Please describe, at a high level, some challenges faced in developing your system to this stage of maturity in this functional area.

FUTURE INITIATIVES

If there are any current or planned development efforts on the system towards higher maturity levels, please describe. Include timelines or milestones as available.

SCALABILITY

Scalability concerns the readiness of a system to handle increased data or user volume.

① AD HOC	② REPEATABLE	③ ENHANCED	④ MANAGED	⑤ OPTIMIZED
The system's (largely manual) processes bottleneck even under normal volume	The system has some manual processes, and has limited support for additional volume of data, data sources, or users	The system is reasonably automated, but spikes in volumes and usage require extra monitoring; could add additional data sources if volume is mitigated	The system is entirely automated, but there is reason to doubt the processes in place could handle a large spike in data/usage, or many new sources	Fully automated support for any conceivable data/usage volume and additional sources

SCALABILITY DEFINED BY THEMES

- This row may require a new theme addressing technological sophistication of data stores.
- **Automation** – How robust is the automation of a system In the face of increased data and use?

CURRENT CAPABILITIES

Please rank the system on the above maturity scale. If your system does not cleanly fall into any one category, please select portions of the above to briefly describe the state of your system's data. In either case, please provide detail to support your assessment.

LESSONS LEARNED

Please describe, at a high level, some challenges faced in developing your system to this stage of maturity in this functional area.

FUTURE INITIATIVES

If there are any current or planned development efforts on the system towards higher maturity levels, please describe. Include timelines or milestones as available.

This page intentionally blank.

C. RELATIONSHIP TO AGENCY ARCHITECTURE STANDARDS (RESERVED FOR FUTURE USE)

This appendix is reserved for possible future use.

This page intentionally blank.

D. SERVICE MAPPING TO THE IC JARM (FOUO ONLY)

*A mapping of the IC JARM Service Components and Data Aggregation Services will be available in an FOUO version of the document at a later date.

This page intentionally blank.

E. SAMPLE MISSION USE CASES

The DARA working group intends to address use cases through definition, completion, and piloting of the DARA program. Generally, the possible use cases include:

USE CASE ASSUMPTIONS

All use cases include the following assumptions:

- Organizations have made some information available for query or reference by other agencies' systems in a manner that would provide the analyst an indication of a correlation and some measure of confidence for its accuracy based on a common maturity model.
- A significant discovery could be investigated through channels and methods in accordance with laws, regulations, businesses processes, and security programs applicable to both the data and the analyst.

Organizations implement business processes that enable inter-agency cooperation so that analysts can maximize the value of accessing correlated data through interoperable systems.

USE CASE 1

A user at Organization A is involved in an investigation that involves Entity 1 pertaining to a probable or rapidly emerging terrorist threat. The user performs a query on a correlation system at Organization B to determine if that organization has information on Entity 1 that would further the investigation and enable more rapid interdiction of the threat. The result, for the analyst, could be described first as “yes” or “no” and then, if “yes”, some additional information about the entity.

BENEFIT

The DARA framework makes this inter-organization query possible. Typically the user would have to engage in a significant amount of cross-agency coordination in order to collaborate over an investigation. This coordination can be extremely time or resource intensive and is potentially wasted if, in this case, Organization B did not have any information on Entity 1—a fact that the user would only have known after applying significantly more time and effort. In this way, DARA enables inter-organizational queries that are much faster, leading to faster interdiction, and much less resource-intensive, leading to greater efficiencies.

USE CASE 2

A user at Organization B wishes to identify non-obvious relationships that could precede a terrorist attack but which are not currently known. The user queries multiple systems belonging to different organizations that use the DARA to enable interoperability.

BENEFIT

The resulting interoperability between systems establishes a new layer of data or metadata that can be queried leading to the identification of relationships between resolved entities that could not be identified using just one organization's system. For example, connecting entity information between separate correlating systems may show a single entity that is known in several systems and to several agencies, but the fact that several agencies had identified the entity was not known outside any one agency. When the user discovers that multiple agencies examined the same entity, that information may indicate an emerging threat and warrant further investigation. In this context, the DARA enables identification and investigation of a previously unknown threat. This use case assumes that data developed between interoperable, and connected, systems that follow the DARA could be available for analysis.

USE CASE 3 (CONTRASTING USE CASE)

In today's environment, an engineer at Organization C must receive a bulk data transfer from Organization B on a periodic basis. The bulk data transfer requires significant bandwidth, time to set up, and, inevitably, some time to troubleshoot. Once the data is transferred, the engineer hands the data off to a developer at Organization C who performs additional processing on the data prior to making it available for analysts for queries and investigations. At this point, the data is as old and the periods between data transfers plus the time necessary for scheduling, troubleshooting, and processing have decreased its timeliness and value.

BENEFIT

Organizations that interoperate through the DARA increase the availability of information to analysts by increasing the accessibility of information on other organizations' systems. When analysts are able to search correlated data that other agencies provide using the DARA framework, then organizations do not need to transfer or replicate that information between systems and therefore save storage space, bandwidth, and technical staff members' time across the entire federal enterprise while also enabling searches of data that is more current.

USE CASE 4

Organization E provides a large amount of raw data to Organization F on a regular basis. Organization F's mission requirement is to perform advanced analytics on this data such that

Organization F must have the data in-house. Organization E performs the same data provision services to Organizations G, H, and I, and now must format and standardize the data four different ways to accomplish the transfer. They must also coordinate transfer schedules and SLAs with each of these consumers.

BENEFIT

The DARA framework provides consistent requirements for organizations that provide data to other organizations whose systems interoperate. Rather than meeting requirements that differ by organization or according to differing enterprise architecture frameworks, data providers will have the opportunity to develop requirements against a single reference architecture, lessening time spent on different formatting and standardization processes and errors. Defined services will assist in automating much of the transfer and delivery of these information sets between systems.

USE CASE 5

A user at Organization J has additional information about a resolved entity that they add to the entity record in their organization's system as metadata or additional pieces of information. Systems interoperating using the DARA framework enable immediate access to enriched data as soon as the users query the data or run automated queries, and do not require the interim steps of data transfer and processing.

BENEFIT

This particular use case may be extended as a means for enabling rapid information sharing by, for example, an analyst that identifies a resolved entity in a given organization's context and wants to make the information known to other organizations as it may pertain to investigations or operations in their context. The analyst may add metadata to the resolved entity. This can result in additional contextual information for other analysts that view the entity and lead to greater coordination between organizations and the potential to more quickly convey information about a threat. Note that this additional information will not currently available to users outside the organization until after a bulk-data transfer or load process occurs between organizations.

This page intentionally blank.

F. GLOSSARY OF TERMS

Access Controls: all matters of security policy within a system. This functional area concerns the capability of a system to granularly and interoperably expose subsets of data depending on data's caveats and users' credentials. In the table's wording, both data caveats and user credentials qualify as "access categories"; in assessing the maturity of the system, only the granularity and flexibility with which access categories can be assigned to data should be considered until the Optimized column. To qualify as Optimized, a system should be equally flexible with the degree to which access categories can be assigned to individuals.

Agency: Any executive department, military department, government corporation, government controlled corporation, or other establishment in the executive branch of the Federal government, or any independent regulatory agency.

Attribute: a piece of information or meta-data that determines the properties of a field or tag in a database or a string of characters in a display. These attributes can be used to enable operations, such as search, discovery, or access control, at a level that is more granular that the record-level.

Capabilities: Mission partners and stakeholders have automated computer software-based information systems capabilities that they provide to one another. These capabilities "solve or support a solution for the problems [businesses] face in the course of their business." That is, capabilities are the things organizations have to solve problems and therefore add value, directly or indirectly, to their stakeholders.

Centralized Model: The central, shared entity index model that correlates and disambiguates entities from the indexes made available from within each agency. This central, shared correlation service should utilize a distributed, federated approach to the extent practical, taking into consideration various business needs such as performance.

Change Data Management: The ability of a system to receive and/or propagate data updates and deletions within the system of record from which they originate.

Community: Group of agencies/organizations with a common function, purpose or goal.

Data Correlation: A process for identifying relationships between entities within and across disparate data sets.

Data Aggregation: The collection of processes, policies, procedures, and technologies that allows for the detection of relationships between people, places, things and characteristics, linking information across organizations and helping analysts to identify the connections between data that are not obviously related. A process whereby raw data is gathered and expressed in a summary form for statistical analysis.[33]

Data Aggregation Program: Information technology (IT) capabilities to enable automated correlation of data relating to individuals or entities of interest that might be represented differently in multiple data sets.[34]

Data Federation: Pull summary data from many sources in an attempt to guide other systems and users to source data … Data federation is based on the execution of distributed queries against multiple data sources, federation of query results into virtual views, and consumption of these views by applications, query/reporting tools or other infrastructure components. It can be used to create virtualized and integrated views of data in memory (rather than executing data movement and physically storing integrated views in a target data structure), and provides a layer of abstraction above the physical implementation of data.

Data Harmonization: The process of comparing two or more data component definitions and identifying commonalities among them that warrants their being combined, or harmonized, into a single data component[35].

Data Standard: Agreed-upon structure for representing data in machine-readable format often used to facilitate information exchange through common understanding and recognition of the data elements used[36].

Data Tag: Metadata that helps describe characteristics about the data, such as privacy, security, provenance, source, or other information.

Discovery: The act of locating a description of a Web service-related resource that may have been previously unknown and that meets certain functional criteria. It involves matching a set of functional and other criteria with a set of resource descriptions.[37]

Entity Resolution: The process of determining whether two or more references to real-world objects such as people (individuals), places, or things are referring to the same object or to different objects. This concept is sometimes referred to as Entity Correlation, Entity Disambiguation, or Record Linkage, and includes related concepts such as Identity Resolution. A

[33] http://www-01.ibm.com/software/globalization/terminology/d.html

[34] https://max.omb.gov/community/download/attachments/736986154/2012-0518+ISE+Data+Agg+Capabilities+Report.pdf

[35] https://www.niem.gov/glossary/Pages/Glossary.aspx?alpha=D

[36] https://www.niem.gov/glossary/Pages/Glossary.aspx?alpha=D

[37] https://www.niem.gov/glossary/Pages/Glossary.aspx?alpha=D

set of details that are held about a real-world object such as a person, location, or bank account. An entity is a kind of item.[38]

Entity Map: Complete enriched entity data that includes the linkage of relationships between people, places, things, and characteristics of data resulting from an entity resolution process.

Federated: The process of combining naming systems so that the aggregate system can process composite names that span the naming systems. A relationship in which the participating entities agree to use the same technical standard, enabling access to each other's data and resources.[39]

Guards: Firewalls in place in the network infrastructure to enforce policies related to transfer across domains, including classification domains as well as other categorical domains (i.e., US Person data).

Interoperability: The ability to transfer and use information in a uniform and efficient manner across multiple organizations and information technology systems."[40, 41] It is the ability of two or more systems or components to exchange information and to use the information that has been exchanged.[42]

Master Data Management: MDM is a technology-enabled discipline in which business and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency and accountability of the enterprise's official shared master data assets.

Normalization: The term "normalization" involves transforming data into a common schema that enables the use of a single common data repository.

Message: A message is defined as the entire "package" of information sent between service consumer and service (or vice versa), even if there is a logical partitioning of the message into segments or sections. For instance, if an interface expresses actions as operations or functions that take arguments, and a particular operation has two arguments, both arguments would be considered part of the same message, even though they may be logically separated within the message structure. A message also includes the concept of an "attachment," in which there are several additional sections (attachments) that relate to a distinct, "primary" section.

[38] http://www-01.ibm.com/software/globalization/terminology/d.html

[39] http://www-01.ibm.com/software/globalization/terminology/d.html

[40] Australian Information Interoperability Framework, 2006. http://www.finance.gov.au/files/2012/04/Information_Interoperability_Framework.pdf

[41] United States Code Title 44: Public Printing and Documents (2011) U.S.C. Title 44, Chap. 36, § 3601.

[42] *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries* (New York, NY: 1990).

Mission Speed: In the context of the DARA, Mission Speed is the rate required or desired for enabling access to information so that the need for data does not impede or slow down the mission.

Pattern: A pattern, within the context of this document, is a general, repeatable set of tasks that help accomplish the commonly occurring need for exchange of data or information between two or more exchanging partners.

Probabilistic Records Linkage: Sometimes called fuzzy matching (also probabilistic merging or fuzzy merging in the context of merging of databases), takes a different approach to the record linkage problem by taking into account a wider range of potential identifiers, computing weights for each identifier based on its estimated ability to correctly identify a match or a non-match, and using these weights to calculate the probability that two given records refer to the same entity.

Reference Architecture: Reference Architecture serves as a tool for providing common information, vocabulary, guidance, and direction to guide and constrain architecture and solutions within a particular domain.

Scalability: The readiness of a system to handle increased data or user volume.

Share: Present data for sharing, either within their organization, to another specific organization, or to a community at large with intent to consolidate much of the above functionality for each organization.

Service: A service is the way in which one entity gains access to a capability offered by another entity.

Service Broker: A service broker or intermediary is any capability that receives messages from a consumer and subsequently, as a service consumer itself, interacts with another service. The term "intermediary" indicates that these capabilities sit between other services and "mediate" the interaction by managing, controlling, brokering, or facilitating the transmission of messages between them.

Service Consumer: A service consumer is an entity that seeks to satisfy a particular need through the use of capabilities offered by means of a service.

Service Contract: A service contract is comprised of one or more published documents (called service description documents) that express meta-information about a service. The fundamental part of a service contract consists of the service description documents that express its technical interface. These form the technical service contract which essentially establishes an API into the functionality offered by the service. A service contract can be further comprised of humanreadable documents, such as a Service Level Agreement (SLA) that describes additional quality-of-service features, behaviors, and limitations.[43]

Service Interface: A service interface "is the means for interacting with a service. It includes the specific protocols, commands, and information exchange by which actions are initiated [on the service]." A service interface is what a system designer or implementer (programmer) uses to design or build executable software that interacts with the service. That is, the service interface represents the "how" of interaction.

Service Provider: A service provider is an entity (person or organization) that offers the use of capabilities by means of a service.

Shared Correlated Data: Correlated entity maps, as described in Section 4.1.5.

Shared Raw Data: Some data may be shared in its raw, uncorrelated form. Again, there will ideally be less data exclusively in this category as the environment matures.

Structural Metadata: Considers the flexibility of data's structure. In assessing this functional area, we consider the interoperability of the data formatting standards used and the level of automation with which those standards are implemented and enforced. While metadata functionality is leveraged by many other functional areas (access controls, for instance), the goal of development in the Structural Metadata area is to provide the data scaffolding necessary to support the other functional areas that may leverage it. In making this assessment, please avoid considering factors that depend on the richness of the metadata, but are apart from the structural components of the metadata.

System: A discrete set of information resources organized for the collection, processing, maintenance, transmission, and dissemination of information, in accordance with defined procedures, whether automated or manual.

System Maturity: System Maturity: a benchmark to rate an organization's ability of successfully complete a project based on a hierarchical states of management maturity, management attributes that are critical to the success of any program or organizational endeavor, and the readiness of the technology (materials, components, devices, standards, etc.) used in the project.

Text Extraction: The process of capturing specific text from a document, web page or other text source for use in other systems, processes, objects, etc.

[43] Summarized from http://serviceorientation.com/soaglossary/service_contract

Transport/Infrastructure: The interoperability and automation of a system's transport mechanisms. To a point, it also includes the physical connectivity of a system.

Unshared Data: It is expected that some data will remain unshared from each participating organization. Ideally, there will be less data exclusively in this category as the sharing environment matures.

Whole-of-Government: For the purposes of this reference architecture: 'Whole of government' denotes public service agencies working across portfolio boundaries to achieve a shared goal and an integrated government response to particular issues. Approaches can be formal and informal.

G. ACRONYMS AND ABBREVIATIONS

API	Application Programming Interface
BRM	Business Reference Model
CHISE	Controlled Homeland Information Sharing Environment
CJA	Command Judge Advocate
CT	Counterterrorism
D/As	Departments and Agencies
DARA	Data Aggregation Reference Architecture
DAWG	Data Aggregation Working Group
DoDAF	Department of Defense Architecture Framework
EA	Enterprise Architecture
FEA	Federal Enterprise Architecture
FEAF	Federal Enterprise Architecture Framework
FICAM	Federal Identity, Credential, and Access Management
FISA	Foreign Intelligence Surveillance Act of 1978
GFIPM	Global Federated Identity and Privilege Management
GRA	Global Reference Architecture
I^2F	ISE Information Interoperability Framework
IC	Intelligence Community
IC CIO	Intelligence Community Chief Information Officer
IC ITE	Intelligence Community Information Technology Enterprise
ICAM	Identity, Credential, and Access Management
IDA	Investigative Data Analytics
IEPD	Information Exchange Package Documentation
IISC	Information Integration Subcommittee
INTERPOL	International Criminal Police Organization
IRTPA	Intelligence Reform and Terrorism Prevention Act of 2004
ISA	Information Sharing Agreement
ISAA	Information Sharing and Access Agreements
ISA IPC	Information Sharing and Access Interagency Policy Committee
ISE	Information Sharing Environment
ISO	International Organization for Standardization
IT	Information Technology
JARM	Joint Architecture Reference Model
LEXS-SR	LEISP Exchange Specification – Search and Retrieval

MDM	Master Data Management
MOA	Memorandum of Agreement
MOU	Memorandum of Understanding
MQ	Message Queue
NCTC	National Counterterrorism Center
NIEM	National Information Exchange Model
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
ODNI	Office of the Director of Intelligence
OMB	Office of Management and Budget
ORCON	Originator Controlled
OSI	Open Systems Interconnection
P/CL	Privacy/Civil Liberties
PII	Personally Identifiable Information
PKI	Public Key Infrastructure
PM-ISE	Program Manager, Information Sharing Environment
PPD	Presidential Policy Directive
RDD	Reconciliation Data Dictionary
RISS	Regional Information Sharing Systems
RISSDES	Regional Information Sharing Systems Data Exchange Specification
SAML	Security Assertion Markup Language
SBU	Sensitive But Unclassified
SCI	Sensitive Compartmented Information
SDO	Standards Development Organization
SLA	Service-level Agreement
SLTT	State, Local, Tribal, and Territorial
SOA	Service-Oriented Architecture
SOAP	Simple Object Access Protocol
TOGAF	The Open Group Architecture Framework
TSA	Transportation Security Administration
TSDB	Terrorist Screening Database
UML	Unified Modeling Language
USG	United States Government
WS*	Web Services Specifications
XACML	Extensible Access Control Markup Language